Fostering expert learning strategies in novices

Anique de Bruin

Cover design: Linde Elise Breunis

Lay-out: Legatron Electronic Publishing, Rotterdam

Printer: PrintPartners Ipskamp, Enschede

ISBN

Fostering Expert Learning Strategies in Novices

Het bevorderen van leerstrategieën van experts bij novieten

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus, Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 30 november 2006 om 16.00 uur

door

Anique Bea Hubertina de Bruin
geboren te Maastricht

Promotiecommissie

Promotor:	Prof.dr. H.G. Schmidt
Overige leden:	Prof.dr. F. Gobet Prof.dr. P.E.H.M. Muris Prof.dr. F.G.W.C. Paas
Copromotor:	Dr. R.M.J.P. Rikers

Contents

Voorwoord

Hoewel als laatste geschreven, wil ik beginnen met het leukste deel van dit proefschrift en misschien ook wel het voornaamste, het dankwoord! In een aantal gevallen wil ik mensen bedanken voor hun directe hulp bij het opzetten en uitvoeren van onderzoek, maar vaker nog om indirecte steun in de categorieën afleiding, relativering, interesse, peptalk, en vooral veel lol. In de meeste gevallen ging het om beide. Zonder de eerste vorm van steun had het proefschrift er waarschijnlijk stukken minder uitgezien, zonder de tweede vorm was het schrijven ervan vooral minder leuk geweest!

Als eerste gaat mijn grote bewondering en dank uit naar Remy en Henk voor de fijne samenwerking. Remy in het bijzonder voor zijn twenty-four-seven beschikbaarheid, zijn bereidheid om alles tot in details te bespreken ongeacht de daarmee gemoeide tijd, en zijn vertrouwdheid met het denken van de schaakexpert die het brainstormen over mogelijk relevante onderzoeksopzetten vergemakkelijkte. Henk in het bijzonder vanwege het in mij gestelde vertrouwen, zowel op onderzoeks- als op onderwijsgebied, maar ook vanwege zijn vermogen om te genieten van het doen van onderzoek, dat altijd erg aanstekelijk werkt.

Daarnaast heel veel dank aan Jeroen Bosch, Esther de Kleuver en Simon Gribling van de Nederlandse Schaakbond, voor de fijne samenwerking die heeft geleid tot vruchtbaar onderzoek. Ik hoop dat we in de toekomst onze zoektocht naar verklaringen voor schaakexpertise samen voort kunnen zetten. Ook met stagiaires Jill en Conny was het erg plezierig samen onderzoek op te zetten en uit te voeren, veel dank! Daarnaast veel dank aan Laura, Suzanne, Mirjam, Vera, en Nikky voor hun hulp bij het plannen van de proefpersonen en het verzamelen van de data. Altijd heerlijk wanneer ik druk tikkend achter de computer verrast werd door weer een memory-stick vol met data! Veel dank aan Linde, voor de bijzonder voorkant van dit boekje. Manon, hartstikke veel dank voor het internetdagboek, dat je voor het onderzoek in hoofdstuk 2 en 3 gemaakt hebt.

En natuurlijk dank aan alle collega's van het Instituut voor Psychologie, waarmee het altijd gezellig lunchen, borrelen en dineren is. Bijzonder dank aan mijn voormalige kamergenoten, (in chronologische volgorde) Eric, Wilco, Pascal en Marije, voor het luisterend oor en het eindeloos ouwehoeren. In het bijzonder wil ik ook de O&O-ers en HERC-ers bedanken: Remy, Marije, Susan, Jacqueline, Huib, Peter, Samantha, Wilco, Sofie, en Marieke. Een heel vertrouwde club, met wie het altijd gezellig en gemakkelijk samenwerken is.

Hoewel er inmiddels overlap in namen ontstaat, wil ik toch graag de oude Maastricht club bedanken, met wie we vanaf moment nul vol

overgave eensgezind hebben gewerkt aan het van de grond krijgen van deze psychologie-opleiding met als doel om haar tot de beste van Nederland te maken. Deze groep maakte de overgang van Maastricht naar Rotterdam een peulenschil en stond altijd garant voor bevlogen discussies over onderwijs en onderzoek. Dank Henk, Henk, Marja, Remy, Wilco, Peter en Pascal (en hoewel niet officieel betrokken, hoort Anne hier natuurlijk ook bij). Het was een inspirerende tijd, waar ik met veel plezier aan terugdenk. Ook dank aan de OU-club, in het bijzonder Fred (inmiddels ook 'van ons') en Tamara, voor de interessante discussies over onderzoek en voor de gezellige congresbezoekjes.

In de categorie ontspanning veel dank aan een hoop vrienden (al dan niet georganiseerd in clubjes met obscure namen als Vettige Herders, Tafkatters en Dames 6 van Libanon), voor de reeks oergezellige UNO avondjes, de spontane volleybalweekendjes, de gezellige etentjes, en alle plezier dat nog komen gaat.

Ook dank aan mijn inmiddels officiële schoonfamilie, Peter en Jozien, Rico en Patricia. Fijn om er een tweede gezin bij te hebben.

Heel veel dank aan Boema, voor de oprechte interesse. Ik denk niet dat er veel promovendi zijn die zulke kritische, goede vragen van hun oma krijgen over hun proefschrift, ik voel me zeer vereerd!

Heel veel dank aan mijn broers, Emile en Josse, en aan Dorothé en Miranda. Ondanks onze verschillende levens is het iedere keer dolle pret als we elkaar weer zien.

Enorm veel dank aan mijn lieve ouders, voor de onvoorwaardelijke steun, heel veel liefde en alle goede zorgen. Jullie zijn mijn basis.

Heel veel dank aan Michael en Wilco, dat jullie me vandaag bij willen staan. Maar ook dank voor alle geanimeerde discussies en oprechte interesse voor wat er buiten mijn proefschrift zich afspeelde.

Als laatste natuurlijk veel dank aan Gino, kersverse hubby. Er zijn weinig promovendi die zo intens lief en leed delen tijdens het schrijven van een proefschrift zoals wij dat gedaan hebben, ik vind het heel bijzonder dat je naast het schrijven van je eigen proefschrift ook nog ruimte maakte voor het meebeleven van dat van mij. Maar nog veel meer dank voor alle relativerende ontspanning, ik kijk ontzettend uit naar alle leuke dingen die we samen nog gaan meemaken.

Anique
Rotterdam, september 2006

Chapter 1

General introduction

Fuelled by the English translation of De Groot's seminal dissertation on the cognitive processes of elite chess players (De Groot, 1965), research has, since the end of the 1960s, attempted to characterize defining criteria of expert performance and uncover the cognitive mechanisms that underlie experts' problem solving abilities. The result of these endeavors is an extensive array of studies describing and explaining experts' superior performance in, for example, chess (e.g., Chase & Simon, 1973), computer programming (e.g., Adelson, 1981; McKeithen, Reitman, Reuter, & Hirtle, 1981), sports (e.g., Deakin & Allard, 1991; Starkes, 1987), and medicine (e.g., De Bruin, Schmidt, & Rikers, 2005b; Norman, Brooks, & Allen, 1989). Despite differences in domain and research methodology, these studies replicated the finding that experts were distinguished from non-experts by their near-perfect memory for meaningful stimuli (for a review, see Vicente, 1988; Vicente & Wang, 1998). Theories explaining this phenomenon have generally focused on chess as the domain of study, mainly because of its strong external validity, its complexity that dictates years of experience to achieve expert performance, and its structured nature that allows for clear experimental manipulations (Gobet, 1993). The chunking theory (Chase & Simon, 1973), for example, emphasizes that chess experts not only recall *larger* and *more* units of information. These information units are usually referred to as chunks (Miller, 1956). Whereas non-experts are able to hold up to 7 plus or minus 2 chunks simultaneously in working memory at the same time, Chase and Simon (1973) demonstrated that chess grandmasters had stored hundreds of common constellations of chess pieces as chunks in long-term memory. As these chunks were instantly retrievable from long-term memory, these allowed them to instantly recognize known patterns in a particular chess position. Moreover, having captured more information within a chunk, experts are able to circumvent the working memory limitations non-experts are faced with.

To mention the most prominent theories that followed thereafter, the skilled memory theory (Ericsson & Staszewski, 1989; later expanded as the long-term working memory, Ericsson & Kintsch, 1995) and the template theory (Gobet & Simon, 1996) draw further on the chunking theory to explain experts' superior memory performance. Next to theories that emphasize experts' superior memory for domain information, other studies focus on qualitative differences in the organization of experts' knowledge, and stress that experts' more hierarchically structured knowledge allows them to represent problems in terms of deep theoretical principles, rather than surface features as novices commonly do (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1987; Schmidt, Norman, & Boshuizen, 1990).

The need for deliberate practice to improve performance

Although a detailed comparison of the abovementioned theories is beyond the scope of this dissertation, it is worth noting that they all, either explicitly or implicitly, agree in one respect, which is the necessity of abundant domain-specific experience to acquire expert performance. That is, at this point no documented examples exist of individuals performing at exceptional levels without intense preparation of approximately ten years (Simon & Chase, 1973). Either in the presence or absence of a teacher or coach, the individual striving for excellence is required to dedicate a vast amount of time to training and practice in the domain, be it science, arts, music, or sports. In this regard, Ericsson, Krampe, and Tesch-Römer (1993) maintain that not just any practice, but particularly practice of high quality will foster performance improvement. Their extensive review of studies on learning and skill acquisition led them to conclude that for optimal learning and performance improvement a number of conditions have to be met. First, the individual's motivation to invest effort in the study task is considered indispensable. Next to that, the task should be at a difficulty level that is adapted to the individual's level of preexisting knowledge. If not, the task will be either too easy or too difficult and practice investments will be in vain. Moreover, informative feedback is needed to optimize learning efficiency. Finally, ample opportunity for repetition and correction of errors ensures that suboptimal performance can be improved. In sum, for optimal learning to occur, individuals are required to reflect on their level of performance and regulate further practice behavior by restudying ill-understood parts. Since time is mainly dedicated to rehearsing parts that the individual understands poorly, this type of practice demands a high level of concentration, and will generally not be enjoyable. When practice meets all of the abovementioned criteria, it is termed "deliberate practice." According to Ericsson, Krampe, and Tesch-Römer (1993), deliberate practice is the primary form of time engagement required to foster improvement on the road to excellence.

The first evidence for the deliberate practice assumption came from a comparison of practice behavior of student violinists of several performance levels (from music students becoming teachers to those who were rated as the best by their professors, Ericsson et al., 1993). These students were asked to estimate how many hours per week they had practiced individually for each year since starting to play the violin. The results revealed that weekly practice not only increased steadily since starting age, but also that the accumulated hours of practice differed significantly between the groups of students. The so-called 'best' students had averaged more than 7,000 hours of practice alone, whereas the 'good' group had practiced about 5,300 hours. The cumulative number of hours of practice alone was lowest for the participants studying to become music teachers: 3,420 hours. As practice alone is initiated and directed by the individual, and, therefore, adapted to the individual, it is considered a primary form of deliberate practice in music.

Since their original publication, numerous studies in diverse fields have provided evidence for the tenability of the deliberate practice theory. Not only in chess (e.g., Charness, Krampe, & Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005), but also in sports (e.g., Helsen, Starkes, & Hodges, 1998; Hodge & Deakin, 1998; Hodges, Kerr, Starkes, Weir, & Nananidou, 2004), music (e.g., Sloboda, Davidson, Howe, & Moore, 1996), and medicine (Moulaert, Verwijnen, Rikers, & Scherpbier, 2004) the theory has been tested and confirmed. These studies all related accumulated amounts of deliberate practice to current performance and concluded that a monotonic relation between these two variables was observed.

Fostering expert learning strategies in novices

The theory of deliberate practice is one of the few theories on expertise development that focuses on the learning conditions that foster improvement rather than describing experts' knowledge representations or problem solving behavior (e.g., template theory, or long-term working memory theory). Moreover, although the deliberate practice theory aims at explaining excellent performance, it does not limit its applicability to those who perform at exceptional levels, but mainly describes the intense deliberate practice schedules these individuals have adhered to in order to arrive at that level. Indirectly, this leads to the conclusion that for individuals of all performance levels, deliberate practice will induce performance improvement. In that regard, it would be interesting to study the effect of deliberate practice at the other end of the expertise continuum, that is, in novices. However, until now, research on deliberate practice has mainly concentrated on analyzing practice behavior of individuals who have reached a high level of expertise, and comparing this to individuals of lower expertise. Surprisingly, little research has been carried out studying the effect of deliberate practice prospectively rather than retrospectively, by experimentally comparing the learning gains of a group of novices that practices deliberately to a group that does not practice deliberately (e.g., Shiffrin, 1996). Apart from a study on undergraduates in informal reasoning (Van Gelder, Bissett, & Cumming, 2004), and a study on medical students acquiring ECG interpretation skills (Hatala, Brooks, &

Norman, 2003), no research has examined the effect of deliberate practice on performance in novices. Given the widespread implications this assumption could have on learning in general, it is worthwhile to submit this hypothesis to a closer examination for a number of reasons. First, it can provide insight into the scope and robustness of the deliberate practice theory, and shed light on how performance can be improved, not only in experts, but in groups of more limited skill levels as well. In addition, instead of merely studying the end result of extensive deliberate practice as is traditional in this field, it would enable minute analysis of the developmental process associated with investments in deliberate practice. Finally, research on instruction designed to foster performance in novices tackles a difficult, but highly relevant population, because at some point all learners are novices. Therefore, gaining knowledge about how to stimulate novice learning can lead to the formulation of theoretical and practical implications that will likely find resonance in the educational field.

Despite the external validation for research of this kind, certain theoreticians would question the viability of transfer of expert learning strategies to novices. For example, the expertise reversal effect finds its origin in research demonstrating the inability of novices to benefit from strategies experts use when solving a problem, and vice versa, due to qualitative differences in knowledge organization (Kalyuga, Ayres, Chandler, & Sweller, 2003). Studies demonstrating the expertise reversal effect suggest that instruction designed to foster learning in experts requires adaptation before it can be applied to instruction of low prior-knowledge learners.

Other researchers have argued that low prior-knowledge learners' limited working memory capacity will suffer from expert-like metacognitive activities as self-monitoring and self-regulating learning (e.g., Kanfer & Ackerman, 1989; Winne, 1995a). According to Winne (1995a), these learners will not only commit more errors that require frequent monitoring and self-regulatory activities, but they will also suffer from not having automated these metacognitive processes as expert learners have. Therefore, metacognitive activities place a more frequent and higher demand on low prior-knowledge learners' processing resources than on those of more knowledgeable learners. Winne (1995a) argues that not until learners have at least proceduralized (i.e., automated) the rules of the domain attentional resources become available. Only then will metacognitive activities as self-monitoring and self-regulation demonstrate a beneficial learning effect.

Despite the high working memory demands metacognitive activities entail, others have postulated that in certain, mainly open-ended learning environments (i.e., hypermedia learning), these activities are indispensable to guide learning and foster skill development, even for novice learners (Azevedo, 2005). Moreover, regardless of prior knowledge level, acquiring metacognitive skills contributes to learners' feelings of competence, and will thereby increase motivation to learn (e.g., Brown, 1988; Guthrie, Wigfield, & VonSecker, 2000; Zimmerman & Schunk, 2001). There are, therefore, reasons to assume that learners of low prior-knowledge levels can benefit from expert-like metacognitive strategy training. A set of studies has provided support for this assumption. In these studies, researchers first identified components of the strategic behavior exemplified by experts, and taught these to students, while at the same time having them solve domain problems. In this design, learners concurrently developed domain problem solving skills and metacognitive skills needed to optimize learning. For example, Palincsar and Brown (1984) designed an environment in which learners experienced a set of cognitive strategies in the presence of experts, and gradually came to perform these strategies themselves. Through modeling, the learners were taught four strategies in the domain of reading comprehension (i.e., clarification, question generation, summarization, and prediction). This kind of so-called reciprocal teaching resulted in a strong and positive effect on learning. Similar studies were carried out by Schönfeld (1985) in the domain of mathematics, Larkin and Reif (1976) in the domain of physics, and Scardamalia and Bereiter (1986) in the field of writing. More recently, Butler (1998) showed that providing a training in self-regulatory skills (i.e., the "Strategic Content Learning" approach) to students with learning disabilities led to consistent increases in use of these strategies, and, moreover, to performance improvement on transfer tasks. Finally, Azevedo and Cromley (2004) taught learners to use specific self-regulatory techniques through a 30-minute hypermedia training before studying the circulatory system. Learners who received the training showed a larger shift in their mental model of the circulatory system than those who did not receive the training. This effect could be attributed to extensive use of the self-regulatory processes in the experimental group. In sum, these studies provide evidence that low-knowledge learners, when required to reflect during problem solving, are able to acquire metacognitive skills that eventually prove beneficial to learning.

Although a number of studies have demonstrated the positive effect of metacognitive strategy training in learners with low prior knowledge, some issues have been left unattended. For example, these studies have always selected participants with at least a minimum amount of experience in the domain. We know of no studies that have used participants with no relevant prior domain knowledge whatsoever. Moreover, the strategy training learners received either depended heavily on a proficient teacher who modeled or explained the strategies, or required the development and testing of a complex hypermedia environment. As a consequence, the success of the strategy training varied with the quality of the teacher or the designer of the learning environment. However, in a demanding classroom, a teacher is unable to provide individualized tutoring to students. Therefore, the usability of these trainings is at present debatable. Furthermore, because these strategy trainings were always highly embedded within a specific domain, it is difficult to isolate the elements of the training that most strongly contributed to the learning effect. Therefore, transfer of these elements to other domains in order to meet scientific standards of replicability is virtually impossible, and opportunities for implementation in education are limited.

The studies described in this dissertation were designed to shed light on novices' ability to benefit from metacognitive processes during learning, while taking into account some of the abovementioned shortcomings that were identified in previous research. Our goal was to foster metacognitive activities of novices without relevant prior knowledge of chess through specific experimental manipulations, and study the effect of these activities on learning and transfer performance. Eventually, this could lead to guidelines that describe how to stimulate metacognitive processes in novices and that will possibly find an implementation in educational settings. Because of its relatively high external validity within the field of cognition, and its structured nature that allows unlimited variation in experimental manipulations (Gobet, 1993), we chose chess as the domain of study. Since learning to play chess proficiently is highly complex and requires a vast amount of time, we chose a subpart of the game that could be studied in isolation within the context of an experimental lab setting. This subpart concerned the three-piece theoretical endgame of either King and Rook against King (Chapter 4) or King and Queen against King (Chapters 5 and 6). In these endgames the side that has two pieces (in our studies, the white side) is in theory able to win, but has to apply the correct principles in order to do so. Our goal was to stimulate discovery of these principles by fostering metacognitive processes in participants. In general, the design of the experimental manipulations that intended to trigger metacognitive activities was guided by a number of principles, mostly deduced from previous research in this field (e.g., Boekaerts, 1997; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Thiede & Dunlosky, 1999).

Initially, our goal was to develop an experimental procedure based on the characteristics of deliberate practice and compare its effect to a no-deliberate practice control group. However, because the operationalization of the characteristics of deliberate practice depends on the domain of study and on the population (e.g., what is an adequate level of difficulty for one individual might not be so for the next), and because several studies already exist in educational psychology that tested instructions containing elements of deliberate practice, we decided to broaden our theoretical scope and build upon previous work in this field. Therefore, the design of our experimental procedure will not be described as "deliberate practice", but more generally as instruction that stimulates expert-like metacognitive strategies in novices.

A number of studies have revealed that, when asked to explain to oneself during learning why particular problem-solving steps need to be undertaken, or what the information in a studied text implies, learners' understanding of the problem-solving procedure or of the text improves. This phenomenon has been dubbed the "self-explanation effect" (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). The self-explanation instruction is believed to increase learners' metacognitive awareness, and create opportunities for detection and correction of errors in understanding (VanLehn & Jones, 1993). As this is in line with one of the main assumptions of deliberate practice (Ericsson et al., 1993), the literature on self-explanations provided a relevant theoretical framework for a first study on the effect of metacognitive strategy training in novices. Although the self-explanation effect has been replicated in diverse domains under various circumstances (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, Chiu, & LaVancher, 1994; Neuman & Schwarz, 1998; Renkl, 1997; Renkl, 2002; Renkl, Stark, Gruber, & Mandl, 1998), it has never been examined in novices who have no relevant prior knowledge.

Having examined the effect of generating self-explanations on novice learning, our subsequent studies focused on more specific metacognitive processes in learners. Previous research has shown that theoretical models explaining self-regulated study can be viewed as discrepancy-reduction models (e.g., Butler & Winne, 1995; Dunlosky & Hertzog, 1997; Thiede & Dunlosky, 1999). That is, the individual sets a learning goal, and monitors during study to what extent this goal has been achieved. When a discrepancy is detected between current state of understanding and desired state, the individual determines that restudy is needed. In this approach, two aspects of metacognition are considered essential. The first aspect concerns individuals' self-monitoring of their understanding, whereas the second involves self-regulation of study as exemplified by selection of material that they believe requires restudy. Accuracy of learners' self-monitoring is assessed by correlating the self-produced estimates of understanding with actual performance. A highly positive correlation is indicative of accurate monitoring. In that case, the learner's estimates are consistent with objective performance indicators. These estimates of understanding are usually referred to as "Judgments of Learning" (JOLs; Nelson, Dunlosky, Graf, & Narens, 1990). On the other hand, a measure of self-regulation accuracy is generated when the JOLs of particular pieces of information are correlated with the decision (yes = 1, no = 0) to restudy those pieces of information. A highly negative correlation corresponds with adequate self-regulation. In that case, the individual chose those parts for restudy they believed to have understood poorly. Note that adequate self-regulation depends on accurate self-monitoring: If an individual assigns a high JOL to an item that is not understood, he/she will not select that item for restudy. Self-regulation is in that case poor, and will lead to suboptimal learning gains.

Studies examining these measures of metacognition in relation to learning have revealed that accurate monitoring of understanding and adequate self-regulation of study positively contribute to performance on a posttest (Thiede, 1999). However, studies that demonstrated this effect have mainly used translation pairs as study material (e.g., Swahili-English translations), which from an educational viewpoint have limited relevance. Furthermore, as is the case for the field of self-explanation, research examining the effect of monitoring and self-regulation on learning has disregarded novices as participants. The studies in the final two chapters of this dissertation examined the effect of these more specific metacognitive processes on learning and transfer performance in chess novices.

Overview of the present studies

Before turning towards experimentally testing the learning effect of stimulating metacognitive processes in novices, we first wanted to shed light on the tenability of the theory of deliberate practice in individuals of low prior knowledge in a real-life setting. Therefore, in **Chapter 2** a study is described that analyzes the relation between deliberate practice and performance in a group of elite, adolescent chess players throughout their chess career. Until now, research on the effect of deliberate practice has usually related accumulated amount of deliberate practice to current level of performance (e.g., Charness et al., 1996, 2005; Hodges et al., 2004). Although this approach provides information on the contribution of deliberate practice to performance at a particular point in time, it does not generate insight into the effect of deliberate practice on performance at different developmental stages. For example, does deliberate practice predict performance for those new in the domain, or is the effect merely reserved for those performing at higher expertise levels? Moreover, by only taking into account those who have reached exceptional performance levels, we are not able to deduce to what extent the observed relation between deliberate practice and performance is the direct result of deliberate practice, or rather of some unobserved third variable, such as a talent to practice deliberately. Possibly, the direct cause of experts' exceptional performance does not lie in the excessive number of hours of deliberate practice, but rather in a predisposition to readily engage in deliberate practice. To address this issue, research is needed that compares the relation between deliberate practice and performance of a group of high-performing individuals still active in the domain, to a group of individuals who were promising in their field but eventually dropped out. If both groups benefit from deliberate practice to the same extent, this would suggest that the relation between deliberate practice and performance is not confounded by talent. By means of a linear mixed models analysis approach, we analyzed the relation between chess players' investments in deliberate practice and performance throughout time. Moreover, we compared this relation specifically for dropouts and persistent chess players to assess whether these groups benefit to the same extent from deliberate practice. Finally, the study described in Chapter 2 examined differences in deliberate practice between male and female chess players, and determined its contribution to performance differences between the sexes in this field.

The study reported in **Chapter 3** investigated another until now largely ignored factor in deliberate practice research, namely the influence of achievement motivation on deliberate practice. According to Ericsson et al. (1993) high and sustained motivation is a prerequisite for long-term dedication to deliberate practice. Although research concerning the effect of motivation on time investment in a particular task is prevalent in social psychology, no studies have directly examined the relation between motivation, deliberate practice, and performance. However, such a study could be relevant, as it would provide insight into a possible necessary condition that has to be met in order for deliberate practice to take place. If a high level of motivation turns out to be a prerequisite for deliberate practice, training should not merely be focused at maintaining high levels of deliberate practice, but also at keeping motivation at an optimum. Therefore, the study described in Chapter 3 examined the relation between elite, adolescent chess players' general achievement motivation, motivation to engage in deliberate practice, accumulated hours of deliberate practice, and current chess performance by means of structural equation modeling. Moreover, this study allowed us to analyze to what extent a need for competition contributes to time investment in chess. Until now, research dealing with the influence of competitiveness on intrinsic motivation has shown contradictory findings. Although experimental studies have shown a detrimental effect of competition on motivation, and, hence, on time investments (e.g., Deci, Betley, Kahle, Abrams, & Porac, 1981; Reeve, Olson, & Cole, 1985; Vallerand & Reid, 1984), applied studies in sports have revealed that a competitive orientation might in some cases positively influence time investments (e.g., Brewer, Van Raalte, & Linder, 1993; Gould, Dieffenbach, & Moffett, 2002). The study in Chapter 3 further addresses this issue in the domain of chess.

Whereas Chapters 2 and 3 provide insight into some of the issues that are currently under debate in research on deliberate practice, the studies in Chapters 4, 5, and 6 were directed at examining the effect of fostering metacognitive processes in chess novices on learning and transfer performance. In this regard, the study in **Chapter 4** assessed the effect of providing self-explanations while studying a chess endgame on inference of the endgame principles. Participants were required to study a number of examples of the endgame of King and Rook against King that were played by a chess computer under one of three experimental procedures. The first group of participants was asked to predict the moves the computer was about to make, and simultaneously self-explain aloud why they believed the computer would make that move. The control group was instructed to observe the moves the computer made, without predicting or self-explaining. To isolate the effect of predicting from a possible self-explanation effect, a third group was asked to predict the moves white would make without self-explaining. Learning gains were assessed by analyzing quality and improvement of move predictions over the chess endgames. Transfer performance was determined by requiring learners to play five versions of the endgame against the computer at the end of the experiment and analyze the number of times they were able to checkmate the black King.

Although self-explanation is considered to stimulate self-reflection in learners, the psychological mechanisms that are believed to underlie the self-explanation effect are not entirely metacognitive in nature. For example, part of the effect might be found in learners' more appropriate attendance to the task when self-explaining, a feature that has little connection to higher-level metacognitive processes as self-monitoring and self-regulation. In order to more purely estimate the effect of the latter type of metacognitive processes on novice learning, the studies described in Chapter 5 and 6 tested the effect of stimulating specific metacognitive processes during study of the chess endgame. Therefore, the study described in **Chapter 5** tackles these issues by comparing the learning gains of groups of novice chess players who either monitored understanding by providing JOLs, or who did not monitor understanding. That is, after studying an endgame of King and Queen against King, learners were asked to estimate for every move the computer had to make how confident they were that they would accurately predict a similar move in the future. After this JOL procedure, learners could indicate per move whether they wanted to restudy another example of the move. Monitoring accuracy was determined by correlating JOLs with actual move predictions, adequate self-regulation was indicated by the correlation between JOLs and the decision to restudy the moves (yes = 1, no = 0). To measure the effect of forcing self-regulation upon learners, two groups were asked to select at least two moves per chess endgame for restudy, whereas the other two groups were free in the number of moves they selected for restudy. In a total of four groups, the factor JOL instruction (present or absent) and the factor move selection (free versus forced) were varied. As in the study in Chapter 4, learning gains and transfer performance were assessed.

In the study described in **Chapter 6** the experimental manipulation used in Chapter 5 was adapted, in order to decrease participants' working memory load during execution of the metacognitive activities and thereby improve monitoring accuracy and self-regulation. Moreover, based on the cue-utilization framework (Koriat, 1997), learners were now required to generate relevant cues prior to rating comprehension, in order to further enhance monitoring accuracy. However, before examining the effect of that manipulation, the first experiment reported in Chapter 6 compared the quality of metacognitive activities of inexperienced and more experienced chess players, to determine to what extent these vary with domain experience. Experiment 2 studied the effect of a so-called "generation task" that required learners to activate relevant cues about understanding prior to rating comprehension. The third experiment of Chapter 6 was designed to evaluate whether novice learners can benefit from performance feedback when developing self-monitoring and self-regulation skills. Finally, in **Chapter 7**, a summary and general discussion of the main findings of the studies reported in this thesis are provided, along with reflections on directions for future research.

Chapter 2

Deliberate practice predicts performance throughout time in adolescent chess players and dropouts: A linear mixed models analysis[1]

[1] This chapter is submitted for publication as De Bruin, A.B.H., Smits, N., Rikers, R.M.J.P., & Schmidt, H.G. (2006). *Deliberate practice predicts performance throughout time in adolescent chess players and dropouts: A linear mixed models analysis*. Manuscript submitted for publication.

Abstract

In this study, the longitudinal relation between deliberate practice and performance in chess was examined using a linear mixed models analysis. The practice activities and performance ratings of young elite chess players, who were either in, or had dropped out of the Dutch national chess training, were analyzed since they had started playing chess seriously. The results revealed that deliberate practice (i.e., serious chess study alone and serious chess play) strongly contributed to chess performance. The influence of deliberate practice was not only observable in current performance, but also throughout the chess players' careers. Moreover, although the dropouts' chess ratings developed more slowly over time, both the persistent and dropout chess players benefited to the same extent from investments in deliberate practice. Finally, the effect of gender on chess performance proved to be much smaller than the effect of deliberate practice. This study provides longitudinal support for the monotonic benefits assumption of deliberate practice, by showing that throughout chess players' careers, deliberate practice has a significant effect on performance, and to the same extent for chess players of different ultimate performance levels. The results of this study are not in line with critique raised against the deliberate practice theory that the factors deliberate practice and talent might be confounded.

The necessity of abundant domain-specific experience to acquire exceptional performance has been widely acknowledged in several reviews (e.g., Ericsson, 1996; Ericsson & Lehmann, 1996). Until now, no examples exist of individuals performing at international levels, or improving abruptly without extensive practice in the domain under consideration. In this regard, Simon and Chase (1973) formulated the 10-year rule of expertise, stating that to reach grandmaster level in chess at least ten years of intense preparation are required. Subsequent studies showed that the 10-year rule could be extended to other domains (i.e., mathematics, Gustin, 1985; tennis, Monsaas, 1985; science and arts, Ericsson, Krampe, & Tesch-Römer, 1993). In addition, Ericsson, Krampe, and Tesch-Römer (1993) articulated that especially practice of high quality is beneficial to performance improvement. They argue that 'deliberate practice' is the key activity that determines progress on the road to excellence. Deliberate practice is defined as practice that (1) is primarily directed at performance improvement, (2) is of adequate difficulty, (3) involves informative feedback, and (4) provides ample opportunity for reflection and correction of errors. As such, deliberate practice can be opposed to usual work activities, which do not possess the same degree of control and are generally more directed at problem solving than at performance improvement. Deliberate practice requires full concentration and, therefore, can only be performed for a few hours per day.

To test their theory, Ericsson and colleagues (1993) asked student violinists of several performance levels (from music teacher students to those who were rated as the best by their teachers) to estimate how many hours per week they had practiced alone playing the violin for each year since starting practice. The retrospective practice estimates showed a gradual but marked increase over the years. Moreover, the so-called 'best' students had averaged 7,401 hours of practice alone, whereas the 'good' group had practiced significantly less: 5,301 hours. The cumulative number of hours of practice alone was lowest for the participants studying to become music teachers: 3,420 hours. As practice alone is initiated and directed by the individual, and, therefore, adapted to the individual, it is considered a primary form of deliberate practice.

Since their initial paper (Ericsson et al., 1993), the positive relation between accumulated amount of deliberate practice and performance has been shown in numerous fields as diverse as soccer (Helsen, Starkes, & Hodges, 1998), martial arts (Hodge & Deakin, 1998), triathletes and swimmers (Hodges, Kerr, Starkes, Weir, & Nananidou, 2004), chess (Charness, Krampe, & Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005), music (Sloboda, Davidson, Howe, & Moore, 1996), and teaching (Dunn & Shriner, 1999). In these studies, current expertise level was always monotonically related to accumulated amount of deliberate practice. In contrast, practice measures that were not considered deliberate practice, as tournament play or practicing for fun, did not contribute to current expertise. For example, in chess, Charness, Krampe, and Mayr (1996) analyzed to what extent several practice data and biographic variables, such as start year and club joining age, predicted current chess rating. Their results showed that serious chess study alone, but not analyzing chess games with others or playing in chess tournaments, contributed significantly to current chess rating. Moreover, age at starting to play chess and age at joining a chess club were not significant predictors of current chess rating, as was also the case for having a coach. In a subsequent study, Charness and colleagues (2005) again found that cumulative hours of serious study alone was the strongest predictor of current chess rating. Moreover, total years of private instruction and current hours of serious chess study alone also contributed to performance. However, a possible drawback of their study is that Charness and colleagues failed to evaluate the reliability of the retrospective estimates of practice hours, by comparing them, for example, to self-report diaries (e.g., Ericsson et al., 1993; Sloboda et al., 1996; Starkes, Deakin, Allard, Hodges, & Hayes, 1996).

Monotonic benefits assumption. As noted by Hodges and colleagues (2004), deliberate practice research has mainly focused at predicting differences in performance *between* skill levels, whereas a stronger test would be to examine whether differences in amount of deliberate practice ***within*** skill levels show the same association with performance. Moreover, research on deliberate practice and expertise development has been directed at explaining and predicting current skill level, such as swimming times (Hodges et al., 2004) or ELO ratings in chess (Charness et al., 2005), whereas little research has taken into account the development of performance, and concurrent engagement in deliberate practice, over time. As noted by Hodges and colleagues (2004), research is needed that studies the relation between deliberate practice and performance longitudinally to enable analysis of certain assumptions that can extend the scope of the theory. For example, the deliberate practice theory was initially put forward as a theory explaining performance of top-level experts of national or international level. However, the monotonic benefits assumption, stating that accumulated deliberate practice is monotonically related to current performance, would also imply that within individuals, at any point throughout their career, amount of deliberate practice is associated with performance. The theory does not specify assumptions about delayed effects of practice on performance that would predict zero-order associations when practice and performance are directly related at specific time moments. To test the longitudinal aspect of the monotonic benefits assumption, research is required that studies the relation between deliberate practice and performance, not only for current skill level, but also throughout time.

Given the limited access to longitudinally assessed performance measures, studies addressing the tenability of the monotonic benefits assumption throughout experts' careers are rare. Researchers have mainly approached this issue by relating practice to current skill level. For example, Baker, Côté, and Deakin (2005) compared the performance of three levels of ultra-endurance triathletes (expert, middle of the pack, back of the pack) and related these to accumulated investments in training. While the performance differences between the three expertise levels were equally spaced (i.e., the groups became approximately 2.5 hours slower per triathlon for every lower level of expertise), a different pattern was observed for accumulated hours of training. For instance, the so-called mid-pack and back-pack athletes differed by 2,000 hours of training, whereas the expert and mid-pack athletes differed by 6,000 hours of training. These results suggest that a non-monotonic relation best describes the relation between practice and achievement (Newell & Rosenbloom, 1981). However, because of a failure to include performance measures that were longitudinally assessed, Baker et al. (2005) were unable to study the relation between deliberate practice and achievement throughout career development. Since the monotonic benefits assumption has both theoretical and practical implications that could affect training of non-experts, research is needed that assesses its validity when taking into account the longitudinal development of performance.

A question related to the longitudinal aspect of the monotonic benefits assumption is at what point in career practice is most determining of performance. Until now, research concerning this issue has shown equivocal results. Originally, Krampe and Ericsson (1996) showed that deliberate practice over the last decade was most strongly related to current performance in music. Moreover, according to Baker, Côté, and Abernethy (2003), the effect of practice on performance in sports is delayed, and is not observed until later in the athletes' careers. However, Hodges et al.'s analysis (2004) in swimming and running failed to replicate these findings: an analysis that included the most recent practice years as a predictor of performance explained less variance than a model including recent as well as more distant practice years. Research studying the relation between deliberate practice and performance throughout time can provide further insight into this issue.

Gender differences in cognitive domains. A theme that has emerged more recently in research related to deliberate practice is the influence of gender when explaining variance in performance. Hodges and colleagues (2004) showed that gender, after controlling for differences in deliberate practice, is predominantly a determinant in sports that are of anaerobic nature (e.g., depending on physical factors as muscle strength and distribution of fat in the body), and less so in aerobic activities as distance running or swimming.

However, in the case of cognitive tasks, where muscle strength and fat distribution do not exert an influence, the deliberate practice account does not recognize any genetically based differences between males and females, and asserts that amount of deliberate practice is the foremost determinant of performance.

Although a cognitive task, performance differences between males and females have been widely acknowledged in chess (Howard, 2004). Explanations that have been proposed for this phenomenon concern genetic variation between males and females that could cause specific ability differences (i.e., visuospatial skills) or social factors that lead to distinct socialization environments for boys and girls. While research has shown that males on average tend to score higher than females on tests of visuospatial ability (e.g., Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000), there is little evidence for an association between visuospatial skill and chess performance (Doll & Mayr, 1987; Waters, Gobet, & Leyden, 2002). Frydman and Lynn (1992) showed that young chess players had higher performance IQ scores than non-chess players, but could not prove that this was due to higher scores on measures of visuospatial ability. Therefore, no clear evidence exists for a relation between chess skill and performance on visuospatial tasks, which, in turn, could account for gender differences in chess skill.

The social explanation for gender differences in chess describes how factors as less female role models and fewer opportunities for girls to become high achievers cause lower performance of females (Howard, 2004). Assuming that in the last decades these social factors have diminished considerably, Howard (2004) sets out to explore whether this development is also reflected in a convergence of performance between males and females in chess. To this end, he analyzed the FIDE (World Chess Federation) ratings of all males and females since the start of publication in 1970. However, his data suggest little evidence for a decrease of performance differences between males and females. For example, when looking at the top ranked players, top ranked males still score more than a standard deviation above top ranked females.

There are, however, a number of factors that allow for an alternative explanation of Howard’s results. For example, for the group of players that first appeared on the FIDE list between 1985 and 1989, the males had played 57% more chess games than the females (for the top 100 players, this was even 300%). Moreover, only 39% of the females had an active chess past after 1999, as opposed to 67% of the males, which indicates that the males, for unknown reasons, stayed longer on the list, and probably were more active in chess. Taking this together with the large difference in number of games played, a third explanation not tested by Howard, which concerns a difference in amount of practice, should be further examined. It is possible that males tend to dedicate more time to deliberate practice than females, which could lead to differences in performance. This is substantiated by research from Baker, Côté, and Abernethy (2003), who demonstrated that to become an expert in either basketball, netball, or field hockey, females put in half the number of deliberate practice hours that males put in (2543 hours versus 5159 hours, respectively). Apparently, to compete at an international level in these sports requires less time investment from females than from males. To examine whether a similar explanation applies to chess, research is needed that takes into account both gender and deliberate practice to explain performance.

A further shortcoming in Howard's study is that he combines information from all players that first appeared on the rating list between 1985 and 1989, and he does not control for when people entered or disappeared from the list. In his approach, different background, different entry rating, and a number of uncontrolled social factors might account for part of the variance in ratings. A stronger evaluation of gender differences in career development would be to select a group of chess players of about equal playing strength with similar training opportunities, and to examine whether males and females within that group develop distinctively.

Explaining dropout in promising chess players. Besides the issues of gender and the tenability of the monotonic benefits assumption over time, a fundamental critique raised against the deliberate practice theory is that the factors talent and deliberate practice might be confounded (Sternberg, 1996). That is, because research on deliberate practice has focused on experts who have reached the top, the deliberate practice investments of those who have dropped out along the way have been disregarded. Ignoring those who have not made it, and focusing on those who have, might lead to an overestimation of the relation between deliberate practice and expertise. For instance, if those who dropped out have put in a comparable number of hours of deliberate practice as those who reached the top, this would contradict the suggested monotonic relation between deliberate practice and performance. Perhaps talented individuals use deliberate practice more readily and thus proceed faster on the road to excellence (Sternberg, 1996). To test this hypothesis, research should examine how equal investments in deliberate practice affect performance improvement of dropouts and persisters in chess. If such an analysis reveals that in both groups the monotonic benefits assumption holds, this proves that the often-observed relation between deliberate practice and achievement is not an artifact of subject characteristics (e.g., a predisposition to readily perform deliberate practice). By contrast, this would suggest that regardless of performance level (i.e., persisting or dropping out) equal amounts of deliberate practice improve performance to equal extents.

In sum, although the positive relation between accumulated deliberate practice and performance in high-level experts has been documented in diverse fields, certain issues need to be addressed to examine the scope of the deliberate practice theory. For example, it is at present unclear to what extent the monotonic benefits assumption can be generalized to performance at lower levels, or only applies to performance of high-level experts. If deliberate practice proves beneficial at early stages in career development, this will likely influence non-expert training and teaching. Second, given the lack of controlled studies on gender differences in chess, research is needed that examines the influence of deliberate practice as an explanation for gender differences in chess performance. Finally, the field is in need of research that analyzes the extent to which the effect of deliberate practice is comparable across groups of varying performance success, and not dependent on a predisposition to effectively engage in deliberate practice, by taking into account persistent and dropout chess players.

The purpose of the present study was to test the three assumptions mentioned above. To this end, the relation between (a) deliberate practice (i.e., serious chess study alone and serious chess play, Charness et al., 1996), (b) gender, and (c) the development of chess performance (measured through chess ratings) was examined over time, by using a linear mixed models analysis. While Hodges et al.'s study (2004) provides evidence that within skill levels deliberate practice hours are related to current performance, to our knowledge, no research has studied this relation throughout time. To test these relations, we asked a group of elite, adolescent chess players, who participated in the highest level of national chess training of the Dutch Chess Federation, or players who had participated but at some point dropped out, to complete a questionnaire regarding their chess practice activities since the time they started to play chess seriously. These groups were both qualified for the highest level of chess training provided in the Netherlands. The availability of the chess ratings provided several times a year by the Dutch Chess Federation enabled analysis of the longitudinal relation between chess ratings, deliberate practice, gender, and the decision to quit or proceed seriously playing chess. Because both the deliberate practice estimates and the chess ratings were time dependent, we were able to analyze per time measurement to what extent deliberate practice and performance were related. In contrast to previous deliberate practice studies in chess (Charness et al., 1996, 2005), we assessed the reliability of the deliberate practice estimates by asking a subset of the participants to complete a self-report diary for three consecutive weeks.

Given the monotonic benefits assumption proposed by Ericsson and colleagues (1993) between deliberate practice and performance, we predicted that deliberate practice hours would contribute significantly to chess ratings, irrespective of point in career or time passed since starting to play chess. Furthermore, we predicted that those who were still in the national training would gain higher chess ratings over time, given the higher number of deliberate practice hours they put in, compared to those who had quit. We predicted that a different line of reasoning would apply when taking into account the factor gender. Since no evidence exists that playing chess relies on capacities associated with gender specific advantages, and given that previous research suggests that males and females tend to differ in how much time they dedicate to practice, we predicted that gender would not contribute to performance, once deliberate practice was controlled for.

Method

Sample

A total of 81 adolescent chess players (30 girls, 51 boys) agreed to take part (*M* age = 16.19, *SD* = 2.75, range 12 – 23 years). These chess players were either in the 2003-2004 selection of the national chess training of the Dutch Chess Federation, or were selected in earlier years, but had previously decided to quit (*M* chess rating at time of test = 1944, *SD* = 259). Every year, the Dutch Chess Federation selects about 10 adolescents from the top-performing chess players in the Netherlands, based on chess ratings and information from regional coaches. To ensure that memories of the training years for those who had quit were recent, we only selected the 24 players who had quit between 1999 and 2004. Two of these were not willing to take part. For the remaining group, 18.2% dropped out in 2003, 36.4% in 2002, 13.6% in 2001, 18.2% in 2000, and 9.1% in 1999. Of the 59 participants tested in June 2004 (92.2% of the total national training group at that time), 11 players later decided not to return to the national training after the summer. Therefore, these individuals were in the analyses taken into the group that had dropped out. All in all, the group that still received national training (hereafter referred to as ‘persisters’) consisted of 48 participants (*M* age = 15.13, *SD* = 2.14), whereas the group that had quit the national training during or before the summer of 2004 (hereafter referred to as ‘dropouts’) consisted of 33 participants (*M* age = 17.77, *SD* = 2.85). The age at which the dropouts had quit was on average 16.12 (*SD* = 2.02). This sample is highly representative of the national training group, given that 92.2% of the persisters participated and 91.7% of the dropouts took part. All participants received a small financial compensation at the end of the study.

Materials

Participants completed a paper-and-pencil questionnaire that consisted of four sections. The first two sections are relevant for this study, results from the last two sections are reported elsewhere (De Bruin, Rikers, & Schmidt, in preparation). The first part consisted of biographic information questions (age when learning to play chess, number of chess books, etc.). The second part inquired about players' cumulative hours of 1) serious chess play against other chess players and 2) serious analysis of chess games alone (cf. Charness et al., 1996, 2005). Chess games over the Internet were not taken into account in this measure of serious chess play, because, for this specific sample, these are considered playful interaction.

Apart from the questionnaire, participants' chess ratings were collected with help of the Dutch Chess Federation, from the start of their chess career. The ratings of the Dutch Chess Federation are calculated in the same manner as the ratings by the World Chess Federation. The only difference between Dutch ratings and FIDE ratings is that the former are also based on Dutch tournaments, which are generally not taken into account in the FIDE ratings. The Dutch ratings, therefore, tend to be somewhat lower than the FIDE ratings. However, since we were not interested in the absolute level of the ratings, but in their development over time, we did not transform the Dutch ratings to FIDE ratings. Until 2000, the Dutch Chess Federation published chess ratings twice a year. In 2001, ratings came out three times, and as of 2002 the chess ratings were published four times a year. Depending on when the participants started, two, three, or four measurements are available per year.

Procedure

The persisters filled out the questionnaire individually during a national chess-training weekend in June 2004. Since the dropouts no longer attended the national trainings, a research assistant visited them at home and asked them to complete the questionnaire.

Analysis

The cumulative number of hours spent on serious chess study alone and on serious chess play against others was calculated by multiplying the weekly estimates by 52 and summing the total hours for each year. This probably overestimates the total cumulative hours of chess study alone and serious chess play somewhat, but provides a consistent pattern across individual chess players.

Reliability of retrospective estimates. To maximize reliability of the retrospective estimates, we took a number of precautions:

1. We ensured that participants' memories of their practice activities were as recent as possible. Therefore, chess players who had dropped out of the national training before 1999 were not allowed to participate. To assess possible within-group differences in memory for practice activities, we analyzed whether differences existed in reported number of hours of deliberate practice (sum of total hours of serious chess play, and serious chess study alone) between those who had stopped longer ago (i.e., before the year 2002) to those who stopped in the more recent past (i.e., during, or after the year 2002), using a oneway ANOVA. For both serious chess play and serious chess study alone, no differences between these groups were found, both Fs < 1.
2. Following the encoding specificity principle formulated by Tulving and Thomson (1973), we stimulated participants at the time of retrieving memories of practice to reinstate the mental context at the time of encoding. This technique is also applied in the Cognitive Interview (Geiselman et al., 1984) to increase the accuracy of eyewitness memories, and is by some considered the most effective component of the Cognitive Interview (Memon & Bull, 1991). We asked participants before estimating the number of hours they had spent at a particular age on serious chess play and serious chess study alone, to first attempt to covertly retrieve the following details for themselves, which could act as retrieval cues that would enhance memory for chess playing at that age:
   - What grade were you in? Who was your teacher at that age?
   - Where were you living at that time? What did your room look like?
   - What other hobbies did you have besides chess?
   - How was your free time distributed over the hobbies you had at that time?
   - How did chess fit into your life at that time? Were you in a chess club? How much time did you spend on chess? Were you playing any tournaments?

   Since these questions were intended to stimulate participants' retrieval of relevant memories, it was most natural for them to do so covertly. Therefore, we have no data to present on their answers to these questions.
3. A subsample of the original participants (36 persons: 20 persisters, 16 dropouts, 44.4% of original sample) completed an Internet chess diary for three consecutive weeks, which allowed us to calculate the correlation between retrospective deliberate practice estimates of the last year and actual practice hours as indicated by the diary. Although this technique is commonly applied in deliberate practice research, it has not been used in deliberate practice research in chess (Charness et al., 1996, 2005; Sloboda et al., 1996).

Linear mixed models analysis. In longitudinal data, correlations are typically observed between dependent measurements. To handle this type of data, regular statistical analyses, such as the summary statistic method, do not apply (see Omar, Wright, Turner, & Thompson, 1999, for an evaluation of techniques to analyze repeated measures). Moreover, regular repeated measures analyses of variance typically cannot cope with missing data, and only take into account participants with complete data, who might not be representative of the entire sample. Also, these analyses estimate group effects, and provide no insight into how individuals develop over time. For these reasons, the mixed-effects regression models have become increasingly popular to model longitudinal data. These models include random regression effects that account for the influence of participants on repeated measurements, and thereby enable analysis of individual development over time. Also, in this type of regression analyses, individuals with incomplete data (i.e., due to drop out) can be included in the analysis.

In general, these mixed-effects regression models consist of a within-subjects (level 1) model, and a between-subjects (level 2) model (Laird & Ware, 1982). Consider the following equation that describes the within-subjects model (Greek letters denote population parameters, whereas Arabic letters refer to parameters of an individual):

$$y_{ij} = b_{0i} + b_{1i}T_{ij} + \varepsilon_{ij}$$

This formula describes measurement y of individual i (i = 1, 2,…N participants) on occasion j (j = 1, 2,… n occasions). This model represents the influence of the independent variable time (denoted T) on the outcome variable y for individual i. This outcome is further affected by his/her initial level b_{0i} and an error term $\varepsilon_{ij.}$ The level 2 (between-subjects model) would look like

$$b_{0i} = \beta_0 + v_{0i}$$
$$b_{1i} = \beta_1 + v_{1i}$$

This model indicates that the individual i's outcome is determined by the population initial level β_0 plus a unique contribution for individual i, namely v_{0i}. In addition, in the general model, the individual growth is determined by both the population increase (β_1) plus the unique contribution for individual i (v_{1i}). Combining the level 1 and 2 models, the complete model with random intercept and random slope is formulated as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_{1i}T_{ij} + v_{1i}T_{ij} + \varepsilon_{ij}$$

This model assumes that the variances of the errors are normally distributed. Visual, exploratory inspection of our data (not shown) revealed that individuals mainly varied in start chess rating (i.e., the intercept) and less in growth rate. For instance, for all participants, chess ratings increased at a similar rate over time. Therefore, we decided to apply a fixed slope and only estimate a random intercept. Consequently, in the models that are used in the present study, the unique contribution v_{1i} will not be modelled, i.e., the individual slope (b_{1i}) is equal to the population slope (β_1). To examine differences between persisters and dropouts in growth rate, we incorporated the factor persistence (1 = persisters, 0 = dropouts) as an independent variable in the analysis. We tested several models to examine which variables should be included to provide the best fit to the data. To assess whether a predictor that was added to the model significantly increased the explained variance, we used the log likelihood ratio test, and compared the difference in log likelihood between the model with and the model without the added predictor to a Chi square distribution, with the number of added parameters as the degrees of freedom. If the models under comparison were not nested, Aikaike's Information Criterion (AIC) was used to assess the differential fit of the models. AIC adjusts the log likelihood for the number of parameters that are estimated (Burnham & Anderson, 2002). The model with the lowest AIC represents the best fit to the data.

Model A: *Deliberate practice and performance throughout time.* First, we tested a model that included time, and the two time-varying covariates serious chess play, and serious chess study alone as predictors of chess rating.

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij}$$

In this formula, β_0 and β_1 denote the population intercept and slope, whereas v_{0i} indicates the individual variation from the population intercept. The parameter y_{ij} refers to the outcome on measurement y (chess rating) for individual i on occasion j. Variable T_1 refers to time, P_2 refers to serious chess play, and S_3 to serious chess study alone. Time was coded in monthly intervals, starting when players received their first official chess rating. Since chess players were asked to estimate their practice activities *per year*, whereas chess ratings were available two, three, or four times a year, different chess ratings within a year were coupled with one value of serious chess study alone and one value of serious chess play. We preferred this method above asking chess players to estimate practice hours in three or four-month intervals, as this would probably lead to a decrease in accuracy of estimations. For 78 chess players we had information about chess ratings, serious chess study alone, and serious chess play in overlapping time intervals. Therefore, all the tested models were run on 78 participants. As indicators of the fit of the tested models, we report the p-values and the proportion explained variance (R^2). However, when more predictors are entered in the regression equation, the proportion explained variance will always increase to a certain extent, even when some of the variables make the equation less efficient. Therefore, to correct for the number of predictors, we also report the adjusted R^2 (Stevens, 2001).

Model B: *Including persistence and the interaction between persistence and time.* To examine whether the factor persistence (i.e., comparing those who eventually dropped out of the national training to those who persisted) affected chess ratings, we included the variable persistence (persisters = 1, dropouts = 0) in the analysis. This variable was introduced to assess to what extent differences in chess ratings are due to pre-existing differences between those who ultimately drop out, and those who persist. Also, to examine whether the influence of the factor persistence changes over time, we added the interaction between persistence and time as a predictor. The model was formulated as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij}$$

In this model, D indicates whether participants persisted or dropped out of the national training. To assess whether dropouts and persisters benefited differentially from deliberate practice, we ran two extra models that incorporated either the interaction between persistence and serious chess play (Model B2), or persistence and serious chess study alone (Model B3).

Model C: *Effect of gender.* The last factor we were interested in was to what extent males and females differ in chess rating development, after the factor deliberate practice is accounted for. To this end, we entered gender, and the interaction between gender and time in the analysis. The former was added to account for differences in start chess rating, the latter to assess whether males' and females' chess ratings developed differently over time. The tested model can be summarized as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij} + \beta_8 G_{5ij} + \beta_9 G_{5ij} T_{1ij}$$

In this model, G (0 for females, 1 for males) stands for gender. We also ran separate analyses to study the possible differential effect of deliberate practice on gender, by either incorporating the interaction between gender and serious chess play (Model C2), or gender and serious chess study alone (Model C3). To provide the results for all models, we used version 3.0 of the NLME library (an acronym for non-linear mixed effects) for the statistics package R, described in detail by Pinheiro and Bates (2000).

Results

Biographic information

Since the age difference between the persisters and the dropouts was significant, F (1, 79) = 21.51, MSE = 5.76, $p < .001$, $\eta_p^2 = .21$, we entered age as a covariate in the oneway analysis of variance for number of trainers, chess books and CDs. Covariate effects will only be mentioned when significant. While most of the biographic information did not differ between persisters and dropouts (e.g., starting age of playing chess, starting age formal training, both Fs < 1), the total number of chess trainers had was higher for the persisters than for the dropouts, F (1, 79) = 12.93, MSE = 21.02, $p < .01$, $\eta_p^2 = .14$, as was also the case for total number of chess books and CDs, F (1, 79) = 9.85, MSE = 3617.26, $p < .01$, $\eta_p^2 = .12$. For the latter variable, the covariate effect of age was significant, F (1, 79) = 12.14, MSE = 3617.26, $p < .01$, $\eta_p^2 = .14$, indicating that older players owned more chess books and CDs. The means and standard deviations of the biographic variables are presented in Table 1.

Preliminary analyses on chess ratings and deliberate practice

The correlation between accumulated hours of serious chess study alone and most recent chess rating was $r = .45$ ($p < .001$), whereas the correlation between accumulated hours of serious chess play against others and most recent chess rating was $r = .42$ ($p < .001$). Figure 1 represents the mean hours of serious chess play and serious chess study alone for the first ten years of seriously playing chess. These are graphed separately for persisters and dropouts. This graph shows a marked increase in practice hours across time. Moreover, these data indicate that, especially for serious chess study alone, persisters and dropouts differ in the time they spent on practice. This difference appears to grow over time. In addition, the mean chess rating of the persisters was higher than the dropouts' chess rating, F (1, 79) = 5.07, MSE = 30185.19, $p < .05$, $\eta_p^2 = .06$, when controlling for the effect of age (the dropouts were on average older than the persisters). Mean chess ratings of the persisters was 1986 (SD = 266), mean chess rating of the dropouts was 1868 (SD = 212). As to the effect of gender, males had a significantly higher chess rating than females, F (1, 79) = 19.91, MSE = 39247.54, $p < .001$, $\eta^2 = .20$. Males had a mean chess rating of 1864 (SD = 140), whereas females had a mean chess rating of 1661 (SD = 140).

Table 1

Means and standard deviations of biographic variables for persisters and dropouts

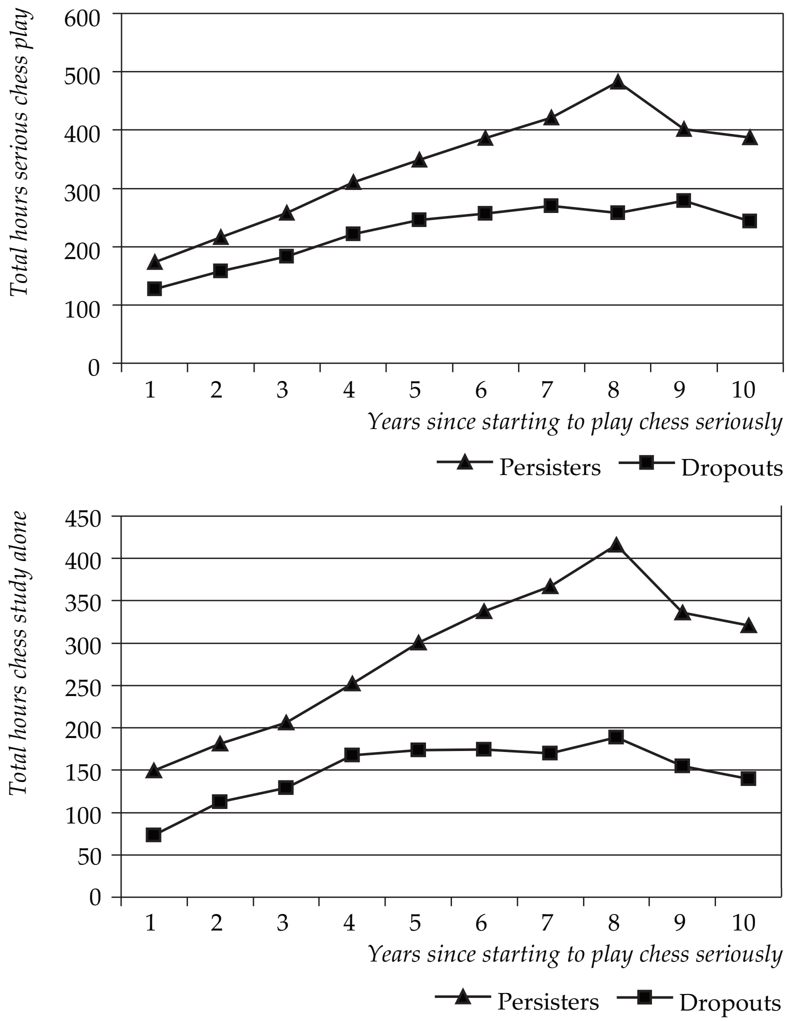
	Persisters		Dropouts		Total	
Question	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age starting to play chess	6.31	1.70	5.80	1.31	6.11	1.56
Age first serious about chess	8.65	2.58	7.94	1.89	8.34	2.33
Age first formal chess instruction	8.10	2.17	8.15	2.23	8.12	2.18
Age first chess rating	11.83	1.29	12.09	1.51	11.94	1.38
Total number of chess trainers had	9.25	5.93	7.24	3.24	8.43**	5.08
Number of chess books	58.17	85.17	31.00	36.27	47.58**	71.21
Number of chess CDs	5.88	4.76	4.20	4.45	5.20**	4.68
Number of chess tournaments played per year[a]	21.01	13.64	21.65	13.46	21.28	13.48
Total number of hours serious chess play	2309.15	1782.19	2263.95	1608.59	2286.55	1702.37
Total number of hours serious chess study alone	1848.45	2415.68	1452.31	1518.35	1650.38	2102.95

[a]It is evident that the persisters nowadays play more chess tournaments than the dropouts. Therefore, we asked the dropouts how many tournaments they played when they still received national training, to be able to compare chess activities during the national training period. That number is provided in the current Table.

*Difference between persisters and dropouts significant at $p < .05$, ** $p < .01$

Figure 1

Number of hours of serious chess play and serious chess study alone per year separated for persisters and dropouts, beginning at the year that they seriously started playing chess



Reliability of retrospective estimates

We calculated the mean weekly hours of serious chess study alone and serious chess play based on three consecutive weeks of diary reports. These means were correlated, using a Pearson product-moment correlation, with the retrospective weekly estimate for these two variables for the current year of practice in the questionnaire. For serious chess play, this correlation was

.74 (N = 36, $p < .01$), whereas, for serious chess study alone, this was .60 (N = 36, $p < .01$). These correlations are comparable to those found in previous studies (e.g., Hodges et al., 2004; Hodges & Starkes, 1996). The lower correlation for serious chess study alone than for serious chess play could be ascribed to the more planned nature of the latter activity. Serious chess play always requires a chess partner, and, therefore, has to be done at specifically planned time periods. Serious analysis alone, however, is directed solely by the individual chess player, and can be done at varying time periods. The variability in moment at which serious chess study alone is performed might lead to somewhat lower memory of these activities compared to serious chess play. For serious chess play, mean weekly retrospective estimate was 5.36 hrs (*SD* = 5.05), whereas the diary mean for this variable was 6.31 hrs (*SD* = 6.10). For serious chess study alone, mean weekly retrospective estimate was 4.38 (*SD* = 6.23), and diary mean was 3.70 (*SD* = 3.79).

Mixed-effects regression models

Model A: *Deliberate practice and performance throughout time.* Model A examined to what extent deliberate practice contributed to chess ratings, when taking into account the time dependency between these variables.

Table 2
Estimated coefficients of model A

Coefficient	Coefficient estimate	Standard error
Fixed parameters		
β_0 (constant)	1575.1826**	20.326
β_1 (T_1, time)	6.778**	0.116
β_2 (P_2, serious play)	0.056**	0.019
β_3 (S_3, serious study)	0.122**	0.020
R^2	0.708	
Adjusted R^2	0.687	
Random parameters		
σ^2_{vi0}	166.063	65.846
Number of parameters	5	
AIC	11207.050	
log-likelihood	-5597.523	

Note. *T* is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. Number of participants = 78, number of measurements = 968. *$p < .05$. **$p < .01$.

In Table 2, the results for the corresponding regression weights (β_0 to β_3) are provided, together with the variance of the random intercept, and the log likelihood of the presented models. This Table shows that all regression weights were significant. The proportion explained variance for this model was high (R^2 = .708).

Model B: *Including persistence and the interaction between persistence and time.* Model B tested whether the chess ratings of the persisters and dropouts developed differently over time. Therefore, the factor persistence and the interaction between persistence and time were entered in the formula. The results showed that the coefficient for the factor persistence was not a significant contributor to the equation, but that the interaction between persistence and time was significant. The positive regression weight for this interaction indicates that persisters' chess ratings improved more than dropouts' ratings. The percentage explained variance for this model was .710. To examine whether addition of the factor persistence and the interaction between persistence and time led to a significantly better fit of the model to the data than Model A, we performed a likelihood ratio test, which resulted in a significant change to the model, χ^2 (2) = 15.61, $p < .0001$. Therefore, we concluded that both the factor persistence and the interaction between persistence and time led to a better fit, and should be included in further analyses.

To study whether persisters and dropouts differentially benefited from serious chess play and serious chess study alone, we entered the interaction between persistence and serious chess play in Model B2, and the interaction between persistence and serious chess study alone in Model B3. For Model B2 (serious chess play), this regression weight was significant. Proportion explained variance for this model was .702. For Model B3 (serious chess study alone), the regression weight was not significant (R^2 = .705). To compare the differential fit of the models B1 through B3, a likelihood ratio test was not possible, because these models were not nested. Instead, we examined the absolute value of the AIC. This comparison showed that the model that included the interaction between persistence and time (Model B1) had the best fit to the data. Apparently, hours of serious chess study alone and serious chess play did not differentially affect the chess ratings of persisters and dropouts. Stated otherwise, both persisters and dropouts benefited to the same extent from serious chess study alone and serious chess play (i.e., deliberate practice). Therefore, we decided to exclude these interactions from further analyses and only incorporate the factor persistence and the interaction between persistence and time.

Table 3

Estimated coefficients of model B and their standard errors

Coefficient	Model B1		Model B2		Model B3	
	Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
Fixed parameters						
β_0 (constant)	1591.98**	31.629	1547.84**	32.533	1566.13**	32.198
β_1 (T_1, time)	6.319**	0.163	6.829**	0.116	6.789**	0.116
β_2 (P_2, serious play)	0.047*	0.019	0.127**	0.030	0.056**	0.019
β_3 (S_3, serious study)	0.110**	0.020	0.128**	0.020	0.150**	0.038
β_4 (D_4, persistence)	-17.834	40.004	43.12	41.849	13.404	41.002
β_5 ($D_4 \times T_1$)	0.905*	0.229				
β_6 ($D_4 \times P_2$)			-0.111**	0.036		
β_7 ($D_4 \times S_3$)					-0.036	0.042
R^2	0.710		0.702		0.705	
Adjusted R^2	0.681		0.672		0.676	
Random parameters						
σ^2_{vi0}	169.276	65.490	170.858	65.665	170.010	66.014
Number of parameters	7		7			7
AIC	11187.42		11197.21			11205.61
log-likelihood	-5585.709		-5590.606			-5594.803

Note. T is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. $D_4 = 0$ for those who eventually dropped out, 1 for those who were still in national training at the time of study. Number of participants = 78, number of measurements = 968. $^*p < .05$. $^{**}p < .01$

Table 4

Means and standard deviations of biographic variables for males and females

Question	Males		Females		Total	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age starting to play chess	6.14	1.65	6.03	1.41	6.11	1.56
Age first serious about chess	8.48	2.35	8.15	2.33	8.34	2.33
Age first formal chess instruction	8.48	2.12	7.53	2.18	8.12	2.18
Age first chess rating	11.57	1.46	12.57	0.97	11.94**	1.38
Total number of chess trainers had	9.20	5.85	7.13	3.06	8.43**	5.08
Number of chess books	64.31	84.68	19.90	21.29	47.58**	71.21
Number of chess CDs	6.39	5.43	3.267	1.98	5.20**	4.68
Number of chess tournaments played per year	19.66	13.26	24.18	13.63	21.28	13.48
Total number of hours serious chess play	2638.17	1901.25	1679.09	1055.63	2286.55*	1702.37
Total number of hours serious chess study alone	2154.85	2477.35	862.74	603.51	1650.38**	2102.95

*Difference between males and females significant at $p < .05$, ** $p < .01$.

Table 5

Estimated coefficients of model C and their standard errors

Coefficient	Model C1		Model C2		Model C3	
	Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
Fixed parameters						
β_0 (constant)	1518.83**	38.638	1485.97**	40.893	1495.39**	39.108
β_1 (T_1, time)	5.855**	0.261	6.329**	0.163	6.317**	0.163
β_2 (P_2, serious play)	0.045*	0.019	0.124*	0.053	0.045*	0.019
β_3 (S_3, serious study)	0.101**	0.019	0.104**	0.020	0.196**	0.061
β_4 (D_4, persistence)	-16.144	37.539	-16.231	37.413	-16.631	41.002
β_5 (D_4 x T_1)	0.930**	0.228	0.902**	0.229	0.915**	0.228
β_8 (G_5, gender)	118.085**	38.397	156.484**	41.342	147.006**	39.221
β_9 (G_5 x T_1)	0.590*	0.265				
β_{10} (G_5 x P_2)			-0.085	0.055		
β_{11} (G_5 x S_3)					-0.098	0.062
R^2	0.749		0.751		0.751	
Adjusted R^2	0.707		0.718		0.718	
Random parameters						
σ^2_{vi0}	158.388	65.336	157.817	65.449	157.772	65.447
Number of parameters	9		9		9	
AIC	11167.02		11172.70		11172.38	
log-likelihood	-5573.509		-5576.350		-5576.190	

Note. T is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. D_4 = 0 for those who eventually dropped out, 1 for those who were still in national training at the time of study. G_5 = 0 for females, 1 for males. Number of participants = 78, number of measurements = 968. *p < .01. **p < .05.

Model C: *Effect of gender.* The last model incorporated the factor gender, and the interaction between gender and time, after the deliberate practice variables, time, and persistence were entered in the analysis. This model had a proportion explained variance of .749. Again, a likelihood ratio test proved that adding these factors led to a significant improvement of model fit, $\chi^2\,(2) = 24.40$, $p < .0001$.

We also examined whether males and females profited differently from serious chess play and serious chess study alone, by adding the interaction between gender and either of these factors in Model C2 and C3, respectively. The AIC indicated that the model with the interaction between gender and time resulted in the best fit to the data. Therefore, the interaction between gender and serious chess play and gender and serious chess study alone were not incorporated in the final model. The final model was of the form:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij} + \beta_8 G_{5ij} + \beta_9 G_{5ij} T_{1ij}$$

This model incorporated the factor time, serious chess play, serious chess study alone, the factor persistence, the interaction between persistence and time, the factor gender, and the interaction between gender and time to account for variance in chess ratings.

Discussion

This study set out to explore the development of the relation between deliberate practice and performance in chess throughout time. We found that serious chess study alone and serious chess play were strong contributors to performance, irrespective of the moment in chess players' careers. By using a mixed models analysis, we were able to explain nearly 75% of the variance in chess ratings throughout young chess players' careers. The variance accounted for solely by the practice-related variables measured longitudinally (71%), was higher than what was previously found in studies on chess (55%, Charness et al., 1996) and music (50-60%, Krampe & Ericsson, 1996). In sum, the mixed models approach used here provided a strong fit to the data, and, moreover, enabled us to study the effect of deliberate practice on performance throughout time.

The current study provides support for the monotonic benefits assumption, which states that current performance is monotonically related to accumulated amount of deliberate practice. Moreover, the findings indicate that, in chess, this assumption not only applies to current performance, but also to the relation between deliberate practice and performance throughout time. That is, regardless of age and performance level of chess players, deliberate practice is monotonically associated with performance throughout chess career. Our findings suggest that, in chess, deliberate practice has an immediate effect on performance, at least when measured at one-year intervals. Although peak performance in chess usually takes place in their mid-thirties, our study shows that from around twelve years onward, chess players' engagement in deliberate practice increases steadily, and that, concurrently, chess ratings improve. These findings emphasize the dedicated and structured manner in which young chess players approach their favourite pastime.

As to the question at what stage in career deliberate practice is determining of performance, our results suggest that, in chess, deliberate practice plays a crucial role very early on, and to the same extent throughout career. These findings diverge from observations in studies in sports, which showed that sport-specific practice was not important until later in athletes' careers (e.g., Baker et al., 2003). However, a difference in setting might explain this discrepancy in results. In sports, individuals usually start by engaging in a wide range of sport activities (i.e., the "sampling years", Côté, Baker, & Abernethy, 2001), before becoming devoted to a single sport (i.e., the "specializing years"). In the sampling years, general motor skills are developed that will prove beneficial to performance in the investment years. Therefore, although practice in the first years is not specific to the sport in which the individual will become an expert, it does have an effect on later performance. As a result, early sport-specific practice is not a necessary condition for expertise, and can be substituted by practice in related sports. By contrast, in chess, domain-specific practice plays a more crucial role and does not have overlap with other activities. Therefore, early chess-specific practice will show a more direct link to chess performance than early sport-specific practice to performance in sports.

Our results revealed that those who dropped out of the national training did not differ a priori from the persisters in chess ratings, but in development of chess ratings over time. That is, both groups had similar chess ratings at the time of selection for the national training, but as time passed, the ratings of the dropouts began to lag behind ratings of the persisters. However, despite the slower growth in chess ratings of the dropouts, both groups benefited to the same extent from deliberate practice, as indicated by the non-significance of the interactions between persistence and serious chess play, and persistence and serious chess study alone. This suggests that the lower chess ratings of the dropouts are not caused by profiting less from investments in deliberate practice than the persisters, but by spending less time on deliberate practice. This is the first study that has tested the critique raised against the deliberate practice theory that the factors talent and deliberate practice could be confounded (Sternberg, 1996). Apparently, in chess, those who are still on the road to expertise are distinguished from those who gave up not by how fast they learn, but by how much time they invest in seriously playing and studying chess. Our results indicate that those who ultimately arrive at expert level in chess do so not because of a predisposition to perform deliberate practice efficiently, but because they put in more hours of deliberate practice. Although all participants were selected for the Dutch national chess training, the variance in chess ratings in this group was considerable, which was best observed in the differences between persisters and dropouts, and males and females, but also in the standard deviations of the chess ratings. Moreover, our sample was young and relatively inexperienced compared to other samples in chess research (e.g., Charness et al., 1996, 2005). This suggests that even within skill levels and regardless of absolute performance, differences in performance could be attributed for more than 70% to variation in time dedicated to deliberate practice throughout career. This finding provides implications for training: Irrespective of skill level, stimulating deliberate practice will likely improve performance. Further research should focus at testing this assumption in other domains besides chess. Moreover, research should study chess players who only recently started playing chess and follow their practice behavior and performance development to provide a prospective examination of this relation.

As to the influence of gender on chess performance, the results reveal that, even after controlling for differences in deliberate practice, males tend to have an advantage in chess ratings over females. However, the non-significance of the interactions between gender and serious chess play, and gender and serious chess study alone indicates that both genders profit equally from deliberate practice. Apparently, similar investments in deliberate practice lead to similar performance improvements for both males and females. The near-significant interaction between gender and time shows that the difference between males and females in chess ratings tends to increase to a certain extent over time. Two remarks are at place here. First, unlike the findings in swimming and running, in our study, adding gender led only to a small increase in explained variance (4% compared to 15% in Hodges et al., 2004). The practice variables, which alone accounted for 71% of the variance, were clearly more important contributors to performance than gender. Second, as our findings indicate that males and females differ to a large extent in hours dedicated to deliberate practice, and in variables as number of chess books and CDs owned, we cannot rule out the possibility that other, not-measured differences in practice-related variables can further explain the performance difference between males and females. Further research is needed that more minutely analyzes the differences in chess activities between males and females. In general, our results provide a more moderate view of the effect of gender on performance compared to Howard's conclusion (2004). We show that, when explaining variation in chess performance, variation in deliberate practice is a much stronger determinant than gender.

In sum, the present study provided new evidence for the tenability of the deliberate practice theory in chess. For the first time, the validity of the monotonic benefits assumption was demonstrated longitudinally, by revealing that groups of different performance development profit similarly from deliberate practice. In future investigations, this assumption needs to be examined in non-cognitive domains. As these results imply that deliberate practice also induces performance improvement in non-experts, further research should focus attention on teaching the characteristics of deliberate practice to individuals of lower expertise. Finally, we detected that gender has a significant, but relatively small effect on performance above the effect of deliberate practice. We encourage use of the mixed models analysis in deliberate practice research in other fields besides chess, in order to extend and refine our knowledge of the effect of deliberate practice on the development of expertise.

Chapter 3

The influence of achievement motivation and chess-specific motivation on deliberate practice[2]

[2] This chapter is submitted for publication as: De Bruin, A.B.H., Rikers, R.M.J.P., & Schmidt, H.G. (2006). *The influence of achievement motivation and chess-specific motivation on deliberate practice.* Manuscript submitted for publication.

Abstract

Although the necessity of high motivation as a precondition for deliberate practice has been acknowledged in the literature, no research has directly tested this relation. Therefore, the present study examined the influence of achievement motivation and chess-specific motivation on accumulated hours of deliberate practice and current performance in a group of elite, adolescent chess players. Moreover, to provide insight into explanations for dropout among promising chess players, the differences in achievement motivation and chess-specific motivation between persistent and dropout chess players were analyzed. The results revealed that mainly competitiveness, but also the will to excel proved relatively strong predictors of investments in deliberate practice. This contradicts findings from previous, primarily experimental research, which suggested that competition had a negative effect on intrinsic motivation. The present results indicate that certain personality features of high performing individuals, such as achievement motivation, might differ from those of their averagely performing peers. The lower achievement motivation of the dropouts in the present study further confirms this. Finally, our results suggest that the motivation to engage in deliberate practice is not solely described as the will to improve performance, but rather as the will to attain exceptional levels of performance.

The positive relationship between accumulated hours of deliberate practice and excellent performance has been shown in a large number of studies in sports (e.g., Baker, Deakin, & Côté, 2005; Duffy, Baluch, & Ericsson, 2004; Helsen, Starkes, & Hodges, 1998; Hodge & Deakin, 1998; Hodges, Kerr, Starkes, Weir, & Nananidou, 2004), chess (Charness, Krampe, & Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005), and music (Sloboda, Davidson, Howe, & Moore, 1996). Typically, these studies assessed the number of hours individuals had spent on deliberate practice, and related these to their current expertise level. According to Ericsson, Krampe, and Tesch-Römer (1993), deliberate practice is defined as practice activities that are primarily directed at performance improvement, and, therefore require an adequate level of difficulty. Moreover, deliberate practice activities involve informative feedback, and provide ample opportunity for reflection and correction of errors. Practice activities that are performed merely for their inherent enjoyment (e.g., practicing for fun), are not considered deliberate, and, as such, do not foster progress on the road to excellence.

More recently, research has turned towards identifying conditions that enable deliberate practice. For example, a study by Hodges and colleagues (2004) revealed that in anaerobic sports (e.g., sprint running or sprint swimming), gender mediates the effect of deliberate practice on performance. More specifically, males tended to benefit more from deliberate practice in these sports than females, whereas this difference disappeared in aerobic sports (e.g., long distance running or swimming). Moreover, Baker and colleagues (2005) showed that non-sport specific practice (i.e., not swimming, cycling, or running, but other sports) can also be beneficial to performance improvements of ultra-endurance triathletes. Apparently, certain largely unidentified variables can mediate the relationship between deliberate practice and performance improvements. One of these factors that has been alluded to in the original study (Ericsson et al., 1993), but has since been ignored in deliberate practice research, is the role of motivation. One can imagine that in order to maintain a demanding schedule of daily deliberate practice, often without external rewards, a high level of intrinsic motivation is indispensable. Ericsson and colleagues (1993) acknowledge this by describing that, in the literature, the most cited condition for optimal learning is the motivation of students to attend to a task and improve performance. According to them, individuals are motivated to practice, because practice improves performance. For example, expert performers are often seen to decrease practice intensity off season (i.e., during the summer time for ice skaters). This supports the hypothesis that deliberate practice is not performed because of its inherent enjoyability, but primarily with an instrumental goal, that is, to improve performance.

This view on the role of motivation on deliberate practice corresponds with what is termed a mastery orientation of motivation (e.g., Helmreich & Spence, 1978). In contrast to a competitive orientation, individuals striving towards mastery of a domain are motivated because they are challenged by difficult tasks and have high internal standards of excellence. That is, they are motivated because they want to master a certain task. Instead, individuals with a competitive orientation are motivated by a desire to win competition against others (i.e., Helmreich & Spence, 1978). Research addressing the influence of either of these orientations on motivation has demonstrated the detrimental effect of a competitive orientation on intrinsic motivation in several experimental studies (e.g., Deci, Betley, Kahle, Abrams, & Porac, 1981; Reeve, Olson, & Cole, 1985; Vallerand & Reid, 1984). Moreover, when motivation is led by an external reward (Pritchard, Campbell, & Campbell, 1987), or when the competitive situation is pressuring (Reeve & Deci, 1996), intrinsic motivation further diminishes. In these studies, intrinsic motivation is typically operationalized as the amount of time participants spent during a free choice period on a relevant task after a certain experimental manipulation (Deci, 1971). Similar results have been found in sports; even though sports contain a certain extrinsic component (e.g., prizes, awards, social pressure), amateur athletes report to be directed more by intrinsic than by extrinsic motives (Vallerand, Deci, & Ryan, 1987). Theoretical explanations for these findings ascertain that extrinsic rewards negatively affect intrinsic motivation, because they are used to control people, and, therefore, undermine individuals' sense of self-determination (Deci & Ryan, 1985). That is, individuals loose their feeling of autonomy, which, in turn, causes a decrease in intrinsic motivation and increases chances of dropping out. Related to the self-determination theory (Deci & Ryan, 1985), the achievement goal theory (Ames, 1992) states that a task-involved climate (i.e., a climate that fosters the need to master a task, and make progress) is positively correlated with intrinsic motivation, whereas an ego-involved climate (i.e., the need to demonstrate high ability, or be perceived as competent) is not, or even negatively correlated with intrinsic motivation, and can lead to dropout (Sarrazin, Vallerand, Guillet, Pelletier, & Cury, 2002; Van Yperen & Duda, 1999). These types of climates are influenced by several situational factors, most directly by a coach (Duda, 2001).

In sum, research, mostly experimental in nature, has shown that a competitive, ego-involved task climate diminishes intrinsic motivation, and decreases time-on-task behavior. Translating this to the development of expertise, and the crucial role of extensive time investments in deliberate practice that are needed to improve performance (e.g., Ericsson, 1996; Ericsson & Lehmann, 1996; Simon & Chase, 1973), one would conclude that in order to maintain motivation to adhere to intense deliberate practice schedules, individuals should entertain a mastery orientation towards the task. However, there are reasons to believe that certain competitive aspects may also contribute to keeping up levels of intrinsic motivation in experts. Research on the personality structure of high performing individuals in sports (e.g., Olympic champions) shows that they are characterized by specific traits that are found to a lesser extent in age mates. They are, for example, highly confident of their abilities, exemplify mental toughness, and are very competitive (Brewer, Van Raalte, & Linder, 1993; Gould, Dieffenbach, & Moffett, 2002). Franken, Hill, and Kierstead (1994) added that, in a group of university students, sport interest was mainly predicted by the motivation to perform high, but also by scores on a measure of competitiveness. One can imagine that in physical sports or non-physical domains as chess, where direct competition and performance feedback are omnipresent, an interest in competition is a prerequisite to remain motivated. All in all, there are reasons to assume that, in high-performing individuals, the often-reported detrimental effect of competition on intrinsic motivation is less likely and might even be reversed.

Related to the issue of the possible need for competition of high-performing individuals, is the question of the personality structure of those who are promising in their field, but eventually drop out. A few studies have shown that ego-involved climates cause lower perceived competence, feelings of less autonomy, and low support by their coach in dropouts (Guillet, Sarrazin, Carpenter, Trouilloud, & Cury, 2001; Pelletier, Fortier, Vallerand, & Brière, 2001). In contrast, those who persist demonstrate high self-determination motivation (Vallerand & Bissonnette, 1992). In these studies, the causes of dropout are assumed to lie in the ego-involved climate that characterizes the training environment. However, personality features, such as differences in competitiveness, are not taken into account in these analyses. An alternative explanation for dropout could be that the dropouts differ from those who persist in how much they strive towards and are motivated by competition. To test this hypothesis, a study is needed that controls for differences in training environment that might be either ego- or task-involving, and that examines differences in aspects of achievement motivation between persisters and dropouts.

This study pursued three goals. Given the absence of research that relates motivation to deliberate practice, we examined the effect of achievement motivation, as a fairly stable personality trait, and motivation to perform deliberate practice on accumulated hours of deliberate practice and performance in a group of young elite chess players. Achievement motivation was assessed with use of the Work and Family Orientation Questionnaire (WOFO, Helmreich & Spence, 1978; Spence & Helmreich, 1983). This measure views achievement motivation as a multidimensional construct, and contains three scales: work, mastery, and competitiveness. The work scale reflects the willingness to work hard and perform well, whereas the mastery scale measures to what extent individuals are driven by a challenge to perform high on difficult tasks. The competitiveness scale indicates the degree to which individuals look for competition, and strive towards winning. To map participants' chess-specific motivation to engage in deliberate practice, we designed an instrument that consisted of items assessing individuals' need for performance improvement and need for competition in chess. Using structural equation modeling, we examined the structural relations between the WOFO, the deliberate practice motivation questionnaire, accumulated hours of deliberate practice, and performance level, as indicated by current chess rating. We hypothesized that both measures of motivation would positively contribute to investments in deliberate practice, and thus, chess performance. Secondly, given the importance of competition in elite chess players, and previous research on the personality of high performing individuals, we hypothesized that competitiveness by itself would contribute to deliberate practice and chess performance. Finally, in view of the low inherent enjoyability of deliberate practice, we hypothesized that a model that depicted motivation as the result of deliberate practice and chess performance would exemplify a lower fit to the data than a model that described motivation as the cause of deliberate practice and chess performance.

The third goal of this study was to shed further light on the causes of dropout in young promising individuals, and, more specifically, the possible role of certain personality characteristics. We hypothesized that dropouts would demonstrate lower achievement motivation, and lower deliberate practice motivation. Especially achievement motivation is assumed to represent a stable personality construct that, if different between persisters and dropouts, possibly influences the decision to quit. Therefore, we analyzed differences in achievement motivation and chess-specific deliberate practice motivation between a group of young, elite chess players, who were still in the selection of the Dutch national chess training, and a similar group of chess players who had dropped out of the Dutch national chess training.

Method

Sample
A total of 81 (30 girls, 51 boys) adolescent chess players participated (*M* age = 16.19, *SD* = 2.75, range 12 – 23 years). These chess players were either in the 2003-2004 selection of the national chess training of the Dutch Chess Federation (59 participants), or were selected in earlier years, but had at some point decided to quit (24 participants) (*M* chess rating at time of test = 1944, *SD* = 259). The Dutch Chess Federation selects about 10 young adolescents per year from the top-performing chess players in the Netherlands, based on chess ratings and information from regional coaches. Of the 24 chess players who had quit between 1999 and 2004 two were not willing to participate. Of the 59 participants tested in June 2004 (92.2% of the total national training group at that time), 11 persons later decided not to return to the national training after the summer. Therefore, these 11 persons were in the analyses taken into the group that had dropped out. All in all, the group that still received national training (hereafter referred to as 'persisters') consisted of 48 participants (18 females, 30 males, *M* age = 15.13, *SD* = 2.14), whereas the group that had quit the national training during or before the summer of 2004 (hereafter referred to as 'dropouts') consisted of 33 participants (12 females, 21 males, *M* age = 17.77, *SD* = 2.85). The mean age at which the dropouts had quit was 16.12 years (*SD* = 2.02). This sample is highly representative of the national training group, given that 92.2% of the national training participated and 91.7% of the dropouts took part. All participants received a small financial compensation after completion of the study.

Materials
Participants completed a paper-and-pencil questionnaire that contained four sections. The first part consisted of biographic information questions (age when started to play chess, number of chess books owned, etc.). Results from this first section are reported elsewhere (De Bruin, Smits, Rikers, & Schmidt, in preparation). As defined by Charness and colleagues (Charness et al., 1996, 2005), deliberate practice in chess is represented by the accumulated number of hours of serious chess play and serious chess study alone. The latter form of practice includes serious analysis of chess games and chess books or magazines individually. Internet chess games were not taken into account in this measure of serious chess play, because, for this specific sample, these are considered playful interaction.

Part three of the questionnaire consisted of 26 items that assessed participants' motivation to engage in chess activities that are specifically directed at improving performance, being a core characteristic of deliberate practice (Ericsson et al., 1993). The items either described participants' need to improve performance in chess (i.e., "I want to become chess grandmaster"), or participants' need to win in chess (i.e., "To me, winning is the most important thing in chess"). This questionnaire was designed to map participants' motivation to perform deliberate practice, and shall be referred to as the DPMQ (Deliberate Practice Motivation Questionnaire). All items were rated on a scale from 1 (completely disagree) to 5 (completely agree). The reliability of the questionnaire was fairly high (Cronbach's α = .84).

The fourth section of the questionnaire consisted of the Work and Family Orientation Questionnaire (WOFO, Helmreich & Spence, 1978; Spence & Helmreich, 1983). The reliability and validity of the WOFO have been tested and confirmed in several studies in diverse domains such as sensation seeking (Schroth & Lund, 1997), sports (Gill, 1988), and aging (Die, Seelbach, & Sherman, 1987). The Work and Family Orientation Questionnaire assesses four dimensions of achievement motivation: work (the desire to work hard and perform well on a task), mastery (having a preference for challenging tasks to meet internal standards of excellence), competitiveness (the enjoyment of personal competition and the desire to win and be better than others), and personal unconcern (the lack of concern with negative reactions of others). Since Spence and Helmreich (1983) report that the personal unconcern scale is of little value in achievement research, we administered only the first three scales of the questionnaire. These three scales consisted together of 19 items (work: 6, mastery: 8, and competitiveness: 5). All items were rated on a five-point scale (1 = completely disagree, 5 = completely agree). Whereas part three of our questionnaire aimed to measure motivation to engage in deliberate practice in chess, the WOFO is assumed to measure intrinsic achievement motivation as a stable personality trait (Spence & Helmreich, 1983). Therefore, we obtained information on participants' chess-specific motivation (DPMQ) and general achievement motivation (WOFO).

Apart from the questionnaire, participants' most recent chess ratings were collected, with help of the Dutch Chess Federation. For the dropouts, the most recent chess rating is probably not the most representative of their chess career, since it is measured after the peak of their chess performance. Therefore, instead of collecting the most recent chess rating, the chess rating at time of quitting the national training was collected for the dropouts. The Dutch ratings are calculated in the same manner as the World Chess Federation (FIDE) ratings. However, the Dutch ratings are also based on Dutch tournaments, which are generally not taken into account in the FIDE ratings. Therefore, the Dutch ratings tend to be somewhat lower than the FIDE ratings. Since we were not interested in the absolute level of the ratings, but in variation in ratings between individuals, we did not transform the Dutch ratings to FIDE ratings.

Procedure

The persisters completed the questionnaire individually during a national chess-training weekend in June 2004. Since the dropouts no longer attended the national training, a research assistant visited them at home and asked them to fill in the questionnaire.

Analysis

The cumulative number of hours spent on serious chess study alone and on serious play against others was calculated by multiplying the weekly estimates by 52 and summing the total hours for each year up until the time of the study. Since the dropouts probably changed their practice behavior drastically after quitting the national training, which might render comparison with persisters problematic after that time, these practice measures were accumulated until the time of quitting the national training for dropouts. To estimate reliability of the retrospective estimates, a subset of the participants (36 individuals: 20 persisters, 16 dropouts, 44.4% of the original sample) completed an Internet diary about their engagement in chess activities. We calculated the mean weekly hours of serious chess study alone and serious chess play based on three consecutive weeks of diary reports. These means were correlated, using a Pearson product-moment correlation, with the retrospective weekly estimate of serious chess play and serious chess study alone for the current year of practice in the questionnaire. For serious chess play, it was .74 (N = 36, $p < .01$), whereas, for serious chess study alone, it was .60 (N = 36, $p < .01$). These correlations are comparable to those found in previous studies (e.g., Hodges et al., 1996, 2004). The lower correlation coefficient for serious chess study alone than for serious chess play could be ascribed to the more planned nature of the latter activity. Serious chess play always requires a chess partner, and, therefore, has to be done at specifically planned time periods. Serious chess study alone, however, is directed solely by the individual chess player, and can be done at all times. The variability in moment at which serious chess study alone is performed might lead to somewhat lower memory of these activities compared to serious chess play. For serious chess play, mean weekly retrospective estimate was 5.36 hours (SD = 5.05), whereas the diary average for this variable was 6.31 hours (SD = 6.10). For serious chess study alone, mean weekly retrospective estimate was 4.38 (SD = 6.23), and the diary average was 3.70 (SD = 3.79).

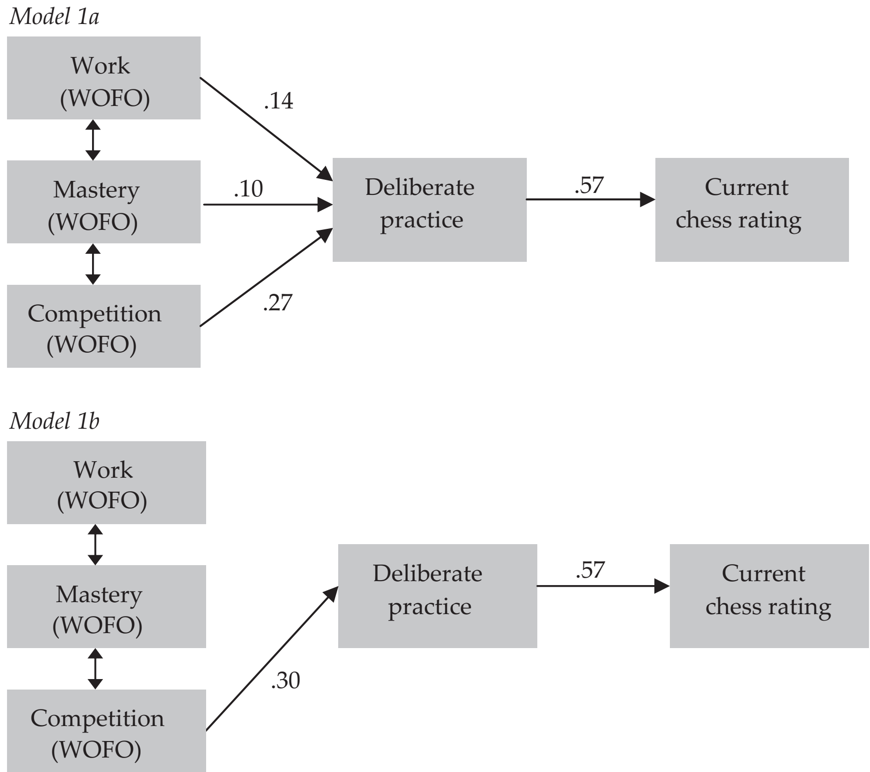
Four types of statistical analyses were performed on the data. First, the correlations between the subscales of the WOFO, and between the DPMQ and the WOFO were calculated. Secondly, the underlying structure of the DPMQ was examined using exploratory factor analysis (principal components estimation with direct oblimin rotation). Afterwards, differences between persisters and dropouts in deliberate practice motivation and achievement motivation, as measured by the DPMQ and the WOFO were computed using oneway analysis of variance. These analyses were run separately on the subscales of these questionnaires as well. Finally, using a structural equation modeling approach, the relations between deliberate practice motivation (DPMQ), achievement motivation (WOFO), deliberate practice, and current chess rating were causally examined (Byrne, 2001). Amos 4.0 was used as statistical program (Arbuckle & Wothke, 1999). Structural equation modeling (SEM) combines multiple regression and path analysis to enable testing of causal relations in a hypothetical model based on the covariance and variance structure in a data set. The advantage of SEM over regular multiple regression is that SEM allows for the use of more than one dependent variable, making the analysis of causal models with more than one structural path between variables possible. AMOS produces several goodness-of-fit criteria indicating how well the tested model accounts for the observed variance structure. The model fit criteria that are commonly used and that were assessed in this study are: 1) the Chi-square divided by the degrees of freedom (CMIN/DF), which is required to be smaller than 3.0 to indicate a reasonable fit of an hypothetical model (i.e., the difference between the hypothetical model and the data should be as small as possible), 2) the Comparative Fit Index (CFI) which compares the fit of the particular model under test with a model in which none of the variables are related. A CFI of .90 or higher indicates that the tested model fits the data well, and 3) The root mean square error of approximation (RMSEA), which represents the root square of the Chi-square divided by the number of degrees of freedom. This value is required to be smaller than .05 to be considered acceptable. To be able to analyze the differential fit of the tested models, we also report the overall Chi-square value.

Several structural models were tested and compared as to their level of fit to the data (See Figure 1). Model 1a analyzed the relationship between the three dimensions of achievement motivation, measured with the WOFO, accumulated hours of deliberate practice, and current chess rating. To study the effect of competitiveness on investments in deliberate practice, and, thereby, on performance, a second model was tested that separately studied the effect of this factor on deliberate practice and chess rating (Model 1b). Given the large body of evidence that describes the monotonic effect of deliberate practice on current performance (e.g., Baker et al., 2005; Charness et al., 1996; Duffy et al., 2004; Helsen et al., 1998; Hodge & Deakin, 1998), we depicted chess rating as the result of deliberate practice in all models.

Since achievement motivation as measured by the WOFO is considered a relatively stable personality trait that is not easily changed by specific situational contexts (Spence & Helmreich, 1983), this measure did not allow us to study a possible reversed effect of performance on motivation that was not included in Model 1. Several studies have shown that, in competitive situations, winning has a positive effect on perceived competence and intrinsic motivation compared to losing (i.e., McAuley & Tammen, 1989; Reeve & Deci, 1996; Vallerand & Reid, 1984). Translated to chess, winning or losing is not only limited to individual chess games, but is more strongly expressed in the three-monthly chess ratings that chess players receive. Based on previous research, this would suggest that chess ratings can have an influence on motivation as well. To test this hypothesis, Model 2 included deliberate practice motivation either as the cause of deliberate practice, and, hence, chess rating (Model 2a), or as the consequence of deliberate practice and chess rating (Model 2b). Although the overall reliability of the deliberate practice motivation questionnaire was high, this is the first time that this questionnaire is applied in research. Therefore, results should be interpreted with caution.

Figure 1

Structural models that examined the relationship between achievement motivation, accumulated hours of deliberate practice, and current chess rating



Finally, Model 3 examined to what extent achievement motivation, as measured by the WOFO, and deliberate practice motivation, as measured by the DPMQ, concurrently contribute to accumulated hours of deliberate practice, and hence, chess rating. Since our goal was to study the relationship between motivation, deliberate practice, and chess performance in general, we did not discriminate the SEM analyses between persisters and dropouts but aggregated their data. Because the deliberate practice data and the chess ratings were collected for all participants during attendance of the national training, differences in these variables after having dropped out cannot have influenced our data.

Figure 2

Structural models that examined the relationship between deliberate practice motivation, accumulated hours of deliberate practice, and current chess rating

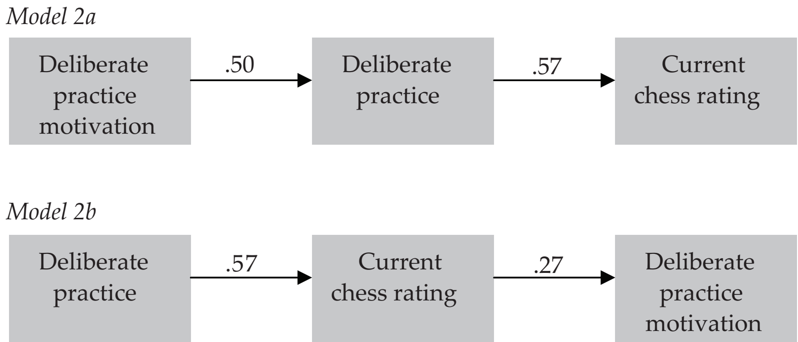
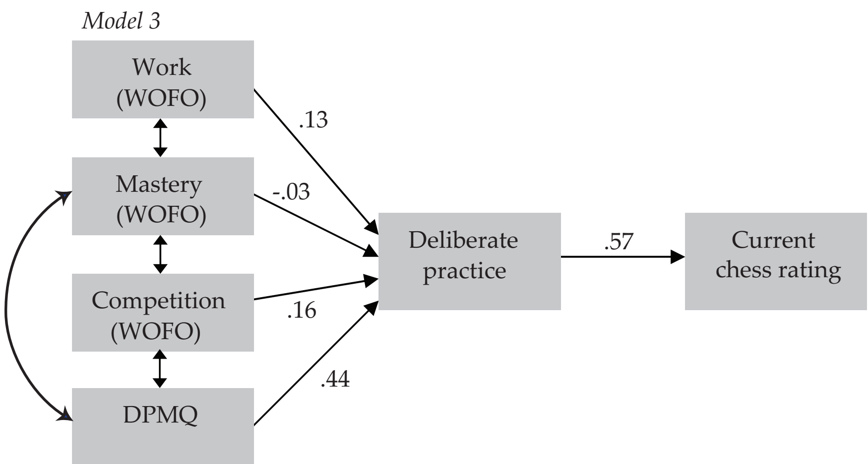


Figure 3

Structural models that examined the relationship between deliberate practice motivation, achievement motivation, accumulated hours of deliberate practice, and current chess rating



Results

Chess ratings and deliberate practice

The correlation between accumulated hours of serious chess study alone and most recent chess rating was $r = .45$ ($p < .001$), whereas the correlation between accumulated hours of serious chess play against others and most recent chess rating was $r = .42$ ($p < .001$). In addition, the mean chess rating of the persisters was higher than the dropouts' chess rating, F (1, 79) = 5.07, $MSE = 30185.19$, $p < .05$, $\eta_p^2 = .06$. Analysis of the development of chess ratings over time and concurrent performance improvement is reported elsewhere (De Bruin, Smits, Rikers, & Schmidt, in preparation).

Validity of the Deliberate Practice Motivation Questionnaire

To gain further insight into the structure of the components underlying these items, an exploratory principal components factor analysis was conducted. Since we suspected a correlation between the measure of performance improvement and the measure of competition, an oblique rotation method was applied (Direct Oblimin, delta = 0). Moreover, because the instrument consisted of 26 items, and therefore the variance explained as represented per eigenvalue is low, the cut off for the eigenvalues was set to 2. These criteria led to the identification of two factors. The first factor had an eigenvalue of 6.51 and explained 25.04% of the variance. The second factor had an eigenvalue of 2.45, and explained 9.42% of the variance. Twenty items predominantly loaded onto the first factor, whereas five items loaded primarily onto the second factor. One item loaded low on both factors ("To improve in chess, analyzing played games is better than playing new games", factor loading 1 = .216, factor loading 2 = .284). One item loaded moderately on both factors ("Improving performance is the most important thing to me when playing chess", factor loading 1 = .389, factor loading 2 = .384). These items were not taken into further analyses. The remainder of the items and their factor loadings are shown in Table 1. Analysis of the content of the items that loaded differently on the two factors led us to conclude that the first factor primarily measured 'the will to excel in chess'. This factor included items as "I want to become a grandmaster", and "I would give up my other hobbies in order to improve in chess if I have to". Clearly, this factor is related to the will to improve performance, but is not identical to it, given the moderate loading of the item "Improving performance is the most important thing to me when playing chess". Instead, the will to excel stretches beyond the will to improve performance, and reflects the desire to reach exceptional levels of performance. The second factor was mainly focused at winning in competition ("My main goal when entering a chess tournament is to gain points for my chess rating", and "To me, winning is the most important thing in chess").

Table 1
Items of the Deliberate Practice Motivation Questionnaire (DPMQ) and their respective factor loadings

Scale	**Item**	**Factor loading**
Will to excel	I want to become a chess grandmaster	.712
	I work on coming up with variations in chess openings	.528
	I would give up my other hobbies in order to improve in chess if I have to	.772
	I want to become a professional chess player	.751
	I've wanted to become a professional chess player for a long time	.634
	I can not imagine ever quitting playing chess	.495
	I think I am able to become a very good chess player	.421
	I want to dedicate as much time to playing chess as possible	.792
	If I would know that I could earn a living by playing chess, I would dedicate more time to it	.579
	I have to play chess every day	.668
	When I start practicing chess, I first stop and think about what I need to improve	.476
	After a played game, I take notes about what I should improve next time	.415
	After a played game, I replay it to analyze what I could have done better	.539
	Even though the chess exercises I make are difficult, I enjoy making them	.439
	I want to be involved in chess in as many ways as possible	.733
	I replay complex parts of a chess game in my head	.374

Table 1 *(Continued)*
Items of the Deliberate Practice Motivation Questionnaire (DPMQ) and their respective factor loadings

Scale	Item	Factor loading
	I prefer playing chess against friends above making the chess exercises for the national training	.346
	When I loose a chess game, I practice harder afterwards	.428
Competition	When I enter a chess tournament, gaining experience is more important to me than winning (mirrored score)	.560
	My main goal when entering a chess tournament is to gain points for my chess rating	.639
	To me, winning is the most important thing in chess	.727
	I learn more from making the chess exercises for the national training than from playing chess against friends	.540
	If I loose against a strong player who taught me a lot when analyzing the game together afterwards, I don't mind having lost (mirrored score)	.524

Differences between persisters and dropouts in motivation
To analyze differences between persisters and dropouts on the two separate factors and on the overall score of the DPMQ, items were summed per factor, and as a whole. First, we analyzed the overall difference between persisters and dropouts on the DPMQ. Persisters scored significantly higher than dropouts on this overall measure, F (1, 79) = 15.99, MSE = 140.25, $p < .001$, $\eta_p^2 = .17$. Separating analyses over the two subscales of the DPMQ, the persisters scored significantly higher than the dropouts on the factor Will to excel, F (1, 79) = 24.97, MSE = 117.66, $p < .001$, $\eta_p^2 = .24$, but not on the Competition scale, F (1, 79) = 1.76, MSE = 117.66, $p = .19$, $\eta_p^2 = .02$. A similar pattern was found for the WOFO, which measures general achievement motivation: dropouts scored significantly lower than persisters, F (1, 79) = 11.18, MSE = 59.92, $p < .01$, $\eta_p^2 = .13$. To examine which subscales contributed to this effect, analyses were performed separately for the three subscales. These analyses revealed that persisters scored significantly higher than dropouts on the mastery scale, F (1, 79) = 4.25, MSE = 17.95, $p < .05$, $\eta_p^2 =$

.05, and that a similar trend was observed for the competitiveness scale, F (1, 79) = 2.79, MSE = 19.89, p = .10, η_p^2 = .03. No effect was found for the work scale, F (1, 79) < 1. The correlation between the WOFO and the DPMQ was relatively high (r = .42, p < .001). Correlations between subscales of these instruments are presented in Table 3.

Table 2

Mean scores on the (subscales of the) WOFO and the Deliberate Practice Motivation Questionnaire (DPMQ). Standard deviations are placed in parentheses.
Note that the number of items, and, therefore, the size of the raw scores, differ between subscales.

	Persisters	**Dropouts**
	WOFO	
Work	23.13 (3.36)	23.12 (2.26)
Mastery	26.49 (4.19)	24.51 (4.31)
Competitiveness	17.53 (4.74)	15.85 (4.01)
Total WOFO	67.15 (8.64)	63.48 (7.48)
	DPMQ	
Will to excel	49.47 (11.97)	37.21 (8.95)
Competition	12.65 (3.70)	13.82 (4.20)
Total DPMQ	64.74 (12.87)	54.03 (10.15)

Table 3

Pearson correlations between the subscales of the Work and Family Orientation Questionnaire (WOFO) and the Deliberate Practice Motivation Questionnaire (DPMQ)

	Subscales of the WOFO		
	Work	*Mastery*	*Competitiveness*
Work	-	.47**	.04 (p = .72)
Mastery	-	-	.22*
	Subscales of the DPMQ		
	Will to excel	*Competition*	
Will to excel	-	-.10 (p = .34)	

Note. * p < .05, ** p < .01.9

Fit of structural models on motivation, deliberate practice, and chess rating

In Figure 1 and 2, the tested models and their corresponding path coefficients are shown. Since this is the first time that the DPMQ is used on a highly representative, but fairly small sample, and because of the skewed distribution of items between the two factors of the DPMQ, we decided to utilize the total DPMQ score instead of the two factor scores in the structural equation analyses. Since preliminary analysis revealed significant correlations between subscales of the WOFO (see Table 3), we decided to enter the correlation between work and mastery, and the correlation between competitiveness and mastery into the model. Given that the correlation between work and competitiveness was zero, this path was not taken into the model.

The path coefficient between accumulated hours of deliberate practice and chess rating was fairly high (.57). This was also found for the relationship between the DPMQ and accumulated hours of deliberate practice (.50). The path coefficients between the scales of the WOFO (Model 1a) and the accumulated hours of deliberate practice were somewhat lower (between .10 and .30). The critical ratios of the path coefficients of the work and mastery scale were 1.142 and 0.792, respectively. Given that these were both smaller than 1.96, this indicates that they were insignificant. Therefore, we decided to test a less constrained model that did not include the paths between work and deliberate practice, and mastery and deliberate practice (Model 1b). Given that the adapted Model 1b is less constrained, the difference in χ^2 should be significant in order to signal a significant improvement of the model fit compared to the more constrained Model 1a. However, the difference in χ^2 between the model with (Model 1a) and the model without (Model 1b) these paths was 3.54 at two degrees of freedom. Since this change in χ^2 is not significant at a cut off score of .05, and since these constructs are theoretically relevant for explaining variation in deliberate practice, we decided to not exclude the work and mastery paths from the final model (Model 1a).

Although the path coefficients do not provide much information about the statistical plausibility of the tested theoretical models, the goodness-of-fit criteria do (Table 4). The goodness-of-fit indices reveal that the model, which takes into account all subscales of the WOFO (Model 1a), provides the best explanation for the data: only for this model is the RMSEA smaller than .05. The model that only relates competitiveness to deliberate practice (Model 1b) also provides a reasonable fit, given the RMSEA of .06. However, as mentioned above, deleting the work and mastery path from the model did not lead to a significant change in χ^2. Apparently, the model that includes all three subscales of the WOFO most adequately reflects the underlying data pattern. The fit of model 1b indicates that this can be largely ascribed to the effect of competitiveness on deliberate practice.

As to the relationship between deliberate practice motivation, deliberate practice, and chess rating, the results reveal the proposed structure: The model that directly relates deliberate practice motivation (DPMQ) to deliberate practice and chess ratings provides a better fit to the data than the model that regards deliberate practice motivation as the result of performance (indicated by chess ratings). This is reflected by both the Chi-square index (Model 2a: $\chi^2/df = 0.04$, Model 2b: $\chi^2/df = 16.59$) and the RMSEA index (Model 2a: RMSEA = 0.00, Model 2b: RMSEA = 0.44). This analysis shows that deliberate practice motivation is best described as the cause rather than the result of deliberate practice and performance.

Given the correlation between the DPMQ and the mastery scale (r = .36, p < .01) and the competition scale (r = .36, p < .01) of the WOFO, we were interested in the concurrent effect of motivation as measured by the WOFO and the DPMQ on deliberate practice, and hence, chess rating. A model including both these instruments allowed us to assess to what extent achievement motivation and deliberate practice motivation are separate constructs. This model contained the relationship between the WOFO (per subscale) and the DPMQ, and, moreover, the correlation between the mastery and the competitiveness subscale of the WOFO on one hand, and the DPMQ on the other hand. The fit of this model (Model 3 in Figure 3) was adequate: The Chi-square divided by the degrees of freedom was 1.03, and the RMSEA was 0.02. A comparison of the path coefficients of this model to the previous models reveals that the subscales of the WOFO now contribute less to variance in deliberate practice, and that variance is largely explained by the DPMQ. Apparently, the more direct measure of deliberate practice motivation more strongly predicts investments in deliberate practice than the measure of the more distant, personality trait achievement motivation. Moreover, the fit of this model underlines the relationship between these two measures of motivation.

Table 4

Fit indices for the tested models on achievement motivation (WOFO), deliberate practice motivation (DPMQ), accumulated hours of deliberate practice, and chess rating

Model name	df	χ^2	χ^2/df	p (χ^2)	CFI	RMSEA
WOFO, deliberate practice, and chess rating						
Model 1a: (WOFO, del. practice, and chess rating)	4	4.13	1.03	0.39	1.00	0.02
Model 1b: (Competitiveness, del. practice, and chess rating)	6	7.67	1.28	0.26	1.00	0.06
DPMQ, deliberate practice, and chess rating						
Model 2a: (DPMQ, del. practice, and chess rating)	1	0.04	0.04	0.85	1.00	0.00
Model 2b: (Del. practice, chess rating, and DPMQ)	1	16.59	16.59	0.00	0.98	0.44
Combined model: WOFO, DPMQ, deliberate practice, and chess rating						
Model 3: (WOFO, DPMQ, del. practice, and chess rating)	6	6.18	1.03	0.40	1.00	0.02

Discussion

This study was conducted to shed light on one of the conditions that is hypothesized to enable deliberate practice, namely, the influence of motivation. Our findings indicate that in a group of elite young chess players, general achievement motivation and chess-specific deliberate practice motivation contributed significantly to variations in accumulated hours of deliberate practice. Moreover, the lack of fit of the reversed model that depicted motivation as the result of deliberate practice and performance supports Ericsson and colleagues' (1993) hypothesis that motivation can be viewed as a precondition for engaging in deliberate practice. Given the higher path coefficient of the relationship between deliberate practice motivation and deliberate practice above the coefficient for the relationship between achievement motivation and deliberate practice, and given the better fit of the former model, our findings suggest that chess-specific motivation possibly more strongly contributes to deliberate practice than achievement motivation. However, given that the model that only included achievement motivation also provided a reasonable fit, this construct in itself does reveal a relevant influence on investments in deliberate practice.

Concerning the conflicting findings in previous research on the influence of competition on motivation, and, therefore, on time on task (e.g., Deci et al., 1981; Reeve et al., 1985; Vallerand & Reid, 1984), our results support the hypothesis that in high-performing, adolescent chess players, competitiveness also contributes to optimizing investments in deliberate practice. Apparently, to remain motivated to adhere to intense deliberate practice schedules in chess, a certain dose of competitiveness is required. Considering the competitive, performance driven climate these chess players find themselves in, this is not surprising. Not only are chess players confronted with winning or losing after every played game, but they are also provided with performance feedback every three months by receiving an updated chess rating, which they can compare against the ratings of their peers. Under these circumstances, those who are not motivated to do their utmost and dislike competition will eventually decide to quit. This line of reasoning is in part supported in the present study: the chess dropouts demonstrated lower achievement motivation than the persisters, which could be attributed to a lower mastery orientation, and a trend in lower competitiveness. Moreover, the dropouts showed a lower will to excel on the deliberate practice motivation questionnaire. Our study reveals that besides an ego-involved training environment as was found in previous research, certain personality features as low achievement motivation can also increase chances of dropping out. Given that achievement motivation is considered as a relatively stable personality trait, the lower achievement motivation observed in the dropouts is probably not the result of lower performance, but as shown in the models tested in this study, the cause.

Although this is the first time the deliberate practice motivation questionnaire was used, and its validity was assessed, the results provide information about the nature of elite chess players' motivation to engage in deliberate practice. That is, the underlying factors of the questionnaire suggest that this motivation is not solely explained by the will to improve performance, but rather by the will to reach exceptional levels of performance. Possibly, these individuals are driven more by long-term achievement goals, than by short-term concrete behavior they desire to improve. This suggests that Ericsson's characterization of deliberate practice as pursuing a performance improvement goal (Ericsson et al., 1993, 1996), might be better described as pursuing a goal of attaining exceptional performance. Research in other fields besides chess is needed to determine whether the need for performance improvement as considered essential to deliberate practice can indeed better be characterized as the need to excel.

In sum, this is the first study that provides evidence for the link between achievement motivation and chess-specific motivation, on one hand, and accumulated hours of deliberate practice, on the other hand. In order to maintain high levels of deliberate practice, achievement motivation and chess-specific motivation appear to play a pivotal role. From both a training and a selection perspective, these findings can provide useful implications. For instance, given the conditional effect of motivation on deliberate practice, motivation should be regularly monitored in individuals to optimize levels of deliberate practice. If motivation is seen to decrease, interventions are needed to improve this. Furthermore, if the current results are replicated in future research, measures of achievement motivation and domain-specific motivation (in this case, chess-specific motivation) can be used as part of the selection process for high-level training. That is, if future studies confirm that motivation is a stable predictor of deliberate practice, and, therefore, of performance, motivation can be assessed and used to optimize the selection process. However, before doing so, some issues require further attention. For example, similar analyses in other domains besides chess can provide insight into the question to what extent our results are specific to the domain of chess, where competition and performance feedback are omnipresent. Possibly, in team sports, where individual performance feedback is less direct, competitiveness is less prevalent, and, therefore, less predictive of deliberate practice and performance. Finally, the retrospective nature of our study did not allow for an analysis of the development of motivation and deliberate practice over time. Achievement motivation is generally considered a stable trait, not easily changed over time. However, chess-specific motivation possibly varies over time as the result of several influences, which could ultimately affect performance (see also Deci, 1971). To provide insight into this issue, future research should take into account the longitudinal development of motivation and concurrent investments in deliberate practice.

Chapter 4

The effect of self-explanation and prediction on the development of principled understanding of chess in novices[3]

[3] This chapter will be published as: De Bruin, A.B.H., Rikers, R.M.J.P, & Schmidt, H.G. (in press). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology.*

Abstract

The present study was designed to test the effect of self-explanation and prediction on the development of principled understanding of novices learning to play chess. First-year psychology students, who had no chess experience, first learned the basic rules of chess and were afterwards divided in three conditions. They either observed (control condition), predicted, or predicted and self-explained the moves of the computer playing a chess endgame of King and Rook against King. Finally, in the test phase, participants had to play the endgame against the computer and were required to checkmate the opponent King. Apart from their test performance, the conditions were compared on quality of move predictions in the learning phase. The self-explanation condition showed better understanding of the endgame principles than the two other conditions, as indicated by the move predictions in the learning phase that more often exemplified correct application of chess principles. Moreover, participants in the self-explanation condition more often checkmated the black King in the test phase than participants in the two other conditions. However, no differences emerged between the prediction and observation condition. This study showed that, even for novices, providing self-explanations stimulates the discovery of domain principles of chess.

In a classic study by Chi and colleagues, individual differences in self-generated explanations of students while solving worked-out examples in physics were investigated (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Their results showed that good problem solvers produced more extensive think-aloud protocols, containing more physics related explanations than the poor problem solvers' protocols. Chi and colleagues concluded that the extent to which learners benefited from the physics worked examples depended on how well participants had explained the study materials to themselves. The problem-related verbalizations produced by good problem-solvers were termed self-explanations (Chi et al., 1989). According to De Leeuw and Chi (2003), self-explanations are any statements related to the topic of study that occur when people explain a problem out loud while solving it, or a text while reading it.

The beneficial effect of self-explanation during problem solving was soon replicated and extended in several studies (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, Chiu, & LaVancher, 1994; Neuman & Schwarz, 1998; Renkl, 1997; Renkl, 2002; Renkl, Stark, Gruber, & Mandl, 1998). These studies showed that not only high quality self-explanations fostered learning, but that self-explaining by itself resulted in greater learning gains than merely reading the study material twice. Even when study time was fixed, the self-explanation effect was still found: Learners who had mastered the study material had provided better self-explanations containing more principle-based explanations than learners who did not master the material (Renkl, 1997).

As to the cognitive mechanisms that might mediate learning from self-explanations, several possibilities have gained support in research. First, verbalization during learning requires learners to process the material in a more deliberate way, thereby leading learners to identify and fill knowledge gaps (VanLehn & Jones, 1993). The more dominant explanation at this moment put forward by Chi (2000) states that self-explanation can aid in revising the mental model that learners have of a text or a problem solving procedure. Mental models, although consistent and coherent, often contain erroneous information. Analysis of self-explanation protocols provides evidence that self-explanations support the process of repairing an incorrect but coherent mental model (Calin-Jageman & Horn Ratner, 2005; De Leeuw, 2000). Chi (2000) adds that this effect is only observed in so-called self-explanation inferences, being those self-explanations that go beyond restating the information provided directly in the text or in the problem, and that consist of inference of new knowledge.

A competing hypothesis states that self-explanations are not used to repair existing mental models, but mainly to aid in building a new mental model of a text or of a problem solving procedure (De Leeuw & Chi, 2003).

When building a text representation, the learner combines the information in the text with background knowledge that is retrieved when reading the text to form a correct situation model that integrates text information with relevant prior knowledge (Van Dijk & Kintsch, 1983). However, when certain parts of the text are unclear, for instance, because the text omits explaining how propositions in a particular sentence are related to propositions from previous sentences, the normal process of constructing a situation model fails. At that point, providing self-explanations can aid by filling in the gaps in the text to derive the correct situation model. Calin-Jageman and Horn Ratner (2005) provide evidence for the latter hypothesis in a non-textual task, namely learning an addition strategy in kindergartners. A crucial aspect of acquiring a correct mental model of a problem solving procedure, is reaching what Gelman and Greeno (1989) and Alexander, Murphy, and Woods (1996) termed 'principled understanding' (see also Alexander, 1997; Renkl, 2002; Renkl, Gruber, & Mandl, 1998). According to Gelman and Greeno (1989) expertise in any domain is not possible without thorough understanding of the central principles of that domain. The principles can be seen as the domain-specific strategies that have to be grasped to be able to adequately solve problems within the domain (VanLehn, 1996). For example, in mathematics, the permutation principle illustrates an elementary probability principle that is often used in cognitive skill acquisition research (for a detailed description, see Catrambone, 1994). Knowledge of the permutation principle is necessary to adequately solve advanced probability problems. Another example is the convergence-of-forces principle in Duncker's famous radiation problem (Duncker, 1945). Correct knowledge of a principle consists of a) knowing the specific formula, b) knowing the variables within the formula, and c) knowing in what type of situations the principle applies.

When learners are able to successfully retrieve a principle and apply it to a problem, they are ready to generalize it to different examples (VanLehn, 1996). At that point, learners possess what is called 'principled understanding'. Thus, principled understanding is a necessary condition for transfer of understanding from one domain to another to occur. Principled understanding enables the student to solve problems, but moreover opens the road to future learning activities of all kinds: Students can try to apply the principle to diverse situations, and thereby revise and expand knowledge of the skill in which the principle is applied. Since principled understanding is adaptive and enables transfer to novel situations, the level of principled understanding can be studied by examining performance and error rates on a transfer task. Previous research has shown that analysis of errors can give insight into the process of knowledge acquisition (e.g., Alexander, Pate, Kulikowich, Farrell, & Wright, 1989; Brown & Burton, 1978). Moreover, the degree to which learners are able to spontaneously verbalize principles also provides an indication of their understanding of the principles.

Until now, research has mostly focused on how principled understanding is exemplified during problem solving, whereas less is known about the actual emergence of principled understanding (Gelman & Brenneman, 1994; Greeno, 1992; Zur & Gelman, 2004). A well-structured domain as chess offers the opportunity to shed light on how understanding of principles develops. In chess, a distinction can be made between the basic rules and the principles that are needed to proficiently play the game. The basic rules cover the legal moves the pieces can make and knowledge of capturing, check, checkmate, and stalemate. These rules exist of the factual knowledge that is necessary to learn the principles of the domain, comparable to knowing what mass is in physics to be able to learn the principle of gravity. Without knowledge of these rules, it is impossible to learn the chess principles. The principles describe the procedures necessary to reach certain subgoals in the game of chess, which are all directed at reaching the ultimate goal, namely checkmate. The principles describe *why* the chess pieces should be moved in a certain way. In sum, the rules explain how to *play* a chess game, whereas the principles instruct how to *win* a chess game. Understanding of the chess principles enables players to understand and predict their opponents' moves and outperform them. For example, the principles that underlie the endgame of King and Rook against King include limiting the space of the King as much as possible, which is necessary to reach the subgoal "Forcing the King to the edge of the board." Because of the complexity and diversity of chess, acquiring principled understanding of chess requires extensive practice. Learning to play at grandmaster level generally takes more than ten years (Simon & Chase, 1973).

Studying the development of principled understanding when learners generate self-explanations implies studying the effect of self-explanations in learners with low prior knowledge. In previous research on the self-explanation effect, participants usually possessed a certain amount of prior knowledge on the topic of study. Participants were for example high school or university students who studied a new problem from a topic they had been studying for some time or that required solution procedures that were already familiar to students (e.g., biology, Chi et al., 1994; Ainsworth & Loizou, 2003; mathematics, Mwangi & Sweller, 1998; probability calculation, Renkl, 1997). Renkl and colleagues (Renkl, Atkinson, & Große, 2004) argue that generating self-explanations during learning is particularly helpful for students at the start of the intermediate phase of skill acquisition, and less so for students in either the early or late phases (for a description of the phases of skill acquisition, see VanLehn, 1996). Although self-explanation involves to a large extent the integration of new information with prior knowledge (Chi et al., 1994), other processing characteristics of self-explanations have been identified that might promote understanding when learners do not possess prior knowledge. For example, Crowley and Siegler (1999) have shown that, when learning to play tic-tac-toe, children (kindergartners, first graders, and second graders) benefited from explaining adults' moves. Despite their low prior knowledge of tic-tac-toe and their young age, the results indicated that the explanations made it easier for children to keep track of subgoal execution, which resulted in higher task performance.

Apart from the study by Crowley and Siegler (1999) on children, to our knowledge, no research has focused on the effect of generating self-explanations on learning during the early phase of skill acquisition in adults. From a cognitive psychological and an educational perspective, it is interesting to examine to what extent learners will be able to benefit from providing self-explanations when they have no prior topic knowledge. If we are able to prove that even novices can profit from providing self-explanations during learning, this will provide insight into how we can optimize novice instruction.

Since a chess position does not provide direct information on the underlying chess principles, chess can be used to study the effect of self-explanation on acquiring principled understanding when explicit information about principles is absent. Learners are required to deduce the principles themselves, by studying multiple examples of chess games. Since principled understanding is crucial to develop a correct mental model of a domain, and self-explanation promotes the development of a mental model, then self-explanation should stimulate the development of principled understanding. Also, in the absence of textual information as is the case in chess, self-explanations cannot be mere rereads or rephrasing of textual information (cf. Chi et al., 1994). Instead, since learners are limited to visual information, and explicit information about the chess principles is absent, they are encouraged to formulate what Chi (2000) has called self-explanation inferences (i.e., self-explanations that indicate inference of new knowledge). This type of inferences has proven to positively affect the mental model learners have of the domain. Moreover, when asking learners to predict the moves of the computer, and explain why the computer would make that move, they are formulating so called *why* explanations. Previous research (Calin-Jageman & Horn Ratner, 2005; Chi et al., 1994) has revealed that producing why-type explanations was correlated with gains in knowledge and skill.

The present study was undertaken to investigate the influence of generating self-explanations on the development of principled understanding of chess when prior knowledge is absent. Participants were in an early phase of skill acquisition according to VanLehn's (1996) classification. After studying an introductory presentation covering the basic rules of chess, participants proceeded to the learning phase and had to learn to play the chess endgame of King and Rook against King by observing a set of endgames played by a computer. Participants were assigned to one of three conditions. In the prediction and self-explanation condition, participants were asked to predict the next move that the computer would make for white (i.e., King and Rook), and at the same time verbally self-explain why the computer would make that move. In order to be able to isolate a possible self-explanation effect from an effect of predicting the next move, a second group of learners only predicted the next move of the computer without further verbalizations (cf. Stark, 1998). Finally, the observation condition merely studied the chess play of the computer without predictions or other verbalizations. In the test phase, all participants had to play five examples of the endgame against the computer and checkmate the black King in as few moves as possible.

We hypothesized that participants in the prediction plus self-explanation condition would be better able to predict the next move of the computer and checkmate the King more often in the test exercises than the two other conditions. Moreover, asking learners to predict the next move requires them to attempt to apply their knowledge of the endgame principles. This will stimulate self-monitoring of comprehension of the principles, which could have a positive effect on learning of the endgame principles. Previous research supports this line of reasoning. For example, Crowley and Siegler (1999) showed that asking children to predict the next step in a game of tic-tac-toe and explaining the expert's strategy facilitated generalization of the strategy to novel situations, by helping them to keep track of subgoal execution. Also, Stark (1998) demonstrated that asking learners to anticipate the next solution step when learning the computation of compound interest positively affected performance on later transfer problems. Therefore, we hypothesized that the prediction condition would outperform the observation condition in the test phase.

As to the discovery of the chess principles that underlie the endgame of King and Rook against King, we hypothesized that the same pattern would occur: Since the prediction and self-explanation instruction would stimulate participants' comprehension monitoring most, this condition would discover more principles of the endgame than the two other conditions, whereas the prediction condition would outperform the observation condition. The content of the self-explanation protocols can provide insight into the cognitive mechanisms that mediate the self-explanation effect and can give an indication of the extent to which learners have acquired principled understanding of chess. If the principles are named repeatedly while self-explaining, the learner exemplifies having acquired principled understanding. To assess within-group differences in self-explanations, we performed a median split and compared the self-explanations of high- and low-explainers (cf., Chi, De Leeuw, Chiu, & LaVancher, 1994). We hypothesized that the high-explainers (those who explicated more than the median number of self-explanations) would have high quality self-explanations that contained more indications of principled understanding, and would perform better in the test phase than the low-explainers (those who verbalized less than the median number of self-explanations).

This study was designed to explore a number of research questions about the role of self-explaining in cognitive skill acquisition:

1. Do self-explanation and prediction when learning to play chess foster the development of principled understanding of chess more than prediction alone? To examine this question, differences between the prediction plus self-explanation condition and prediction only condition in both the number of correct predictions and the application of chess principles in the predictions were analyzed.
2. To what extent does move prediction alone, and move prediction combined with self-explanation foster transfer of understanding of the chess principles to novel problems? This question was examined by testing differences between the three conditions in chess playing performance in the test phase. Moreover, the number of chess principles applied in the test phase and the number of errors made by the different conditions were analyzed.
3. What are the differences in type of self-explanations and test performance between high- and low-explainers? To answer this question, participants were separated by a median split on the number of self-explanations they provided (high- versus low-explainers). This research question provides insight into which high-quality self-explanations foster the development of principled understanding most.

Method

Participants

Participants were 45 first-year psychology students from Erasmus University Rotterdam, The Netherlands. They were randomly assigned to one of the three conditions. To prevent differences in prior knowledge and skill, only students who had never played chess before, nor knew any of the basic rules of chess, were allowed to participate. The students participated to fulfill course requirements, and were awarded a small financial compensation.

Materials

At the beginning of the experiment all participants were shown a computerized presentation providing information about the basic rules of chess. Since it was impossible to learn the complete game during the length of one experiment, a subpart of the game that could be studied in isolation was selected. This subpart involved the theoretical endgame of King and Rook against King (KRK endgame). In this endgame the King and Rook (the white pieces in this study) is theoretically always able to checkmate the black King. In the computer presentation the rules regarding the legal moves for King and Rook were visualized, as well as the basic rules of capturing, check, checkmate, and stalemate.

The chess computer program 'Shredder 6' provided chess exercises, which consisted of examples of the KRK endgame. This chess program plays at Grandmaster level (World Chess Federation (FIDE) chess rating of approximately 2700). The chess exercises ranged in difficulty from checkmate in 6 to checkmate in 10 moves. Moreover, a digital chessboard was connected to Shredder 6 for three purposes. First, after the introductory presentation, the experimenter tested the participants' knowledge of the basic chess rules by asking questions about positions presented on the chessboard. Second, for the chess exercises in the learning phase the experimenter moved the pieces on the chessboard suggested by the computer. During the test exercises the digital chessboard registered the moves that the participants made, after which Shredder 6 generated a countermove, played on the board by the experimenter. Move prediction times and self-explanations were recorded by a digital audio recorder.

Procedure

Studying basic chess rules

The experiment was run during three separate sessions, each lasting about one and a half hour. The three sessions were always planned within a one-week interval. In the first session participants studied the computer presentation about the basic chess rules. On the right hand of the screen, a chessboard was shown, and on the left hand of the screen a certain chess rule was formulated. Sometimes, several slides were necessary to illustrate a single rule. After the rule was explained, learners could test their understanding by answering questions about a specific rule and studying the correct answer on the next slide. This presentation contained 45 slides. It is important to note that the general rules of chess explained in this phase can be clearly distinguished from the principles that underlie the endgame: knowing what the legal moves of a King and Rook are, does not tell a learner anything about how to effectively play the KRK endgame. Extensive practice with varied chess games is needed to acquire principled understanding. Participants were allowed to study the presentation in a self-paced way. This took on average about 20 minutes.

Afterwards, the experimenter tested in a question-and-answer format whether the participant had understood the rules. The questions were of the kind: “In this situation, what moves can the Rook make?”, or “In this situation, is the black King checkmated?”. For every part of the basic rules and concepts (legal moves, capturing, check, checkmate, and stalemate) ten questions were asked. When the participant answered a question incorrectly, the experimenter explained the particular rule in a standard format, repeating the same explanation when the participant still answered the next question incorrectly (e.g., “The black King is checkmated, because it is in check and there is no legal move out of check”). The purpose of this phase was to ensure a baseline of understanding of the basic rules before proceeding to the learning phase. Participants were not allowed to proceed to the learning phase of the experiment before being able to answer these questions correctly. All participants were able to do this before they finished the ten questions of each rule. This took on average about 15 minutes.

Learning phase

Participants studied four chess exercises in the first session and six exercises in the second session of the KRK endgame played by Shredder 6. The second session started with a short rehearsal of the basic rules of chess in the same question-and-answer format as in the first session. The chess exercises were designed to enable participants to acquire understanding of the principles that are essential to the KRK endgame.

Participants in the so-called Prediction and Self-Explanation condition (hereafter referred to as PSE condition) were asked to predict every next move the computer would make for white (King and Rook), and at the same time to self-explain out loud during prediction of the move (i.e., "Explain why you think the computer will make that move"). Since we did not want to limit participants' self-explanation activities, they could take as long as they needed to do this. Every time participants predicted an incorrect move, they were also asked to self-explain aloud the discrepancy between their prediction and the computer move (i.e., the evaluation of computer feedback). This instruction was given to stimulate learners to revise whatever incorrect hypotheses they entertained about the principles that underlie the computer moves.

Participants in the Prediction Only condition (PO condition) predicted the next move of the computer, but were not asked to provide verbal self-explanations when generating a prediction or when receiving computer feedback on the correct move. Finally, participants in the Observation condition were merely instructed to observe the moves the computer made in the chess exercises. The experimenter was present at all times to ensure that participants followed the procedure carefully.

To control for time on task, the mean study times participants in the PSE condition needed to predict the next move were used to determine the prediction and study times for the different moves in the PO and Observation condition. All participants were randomly assigned to one of the three conditions before any of the participants were tested.

Test phase

The third session consisted of five new examples of the KRK endgame. Participants were required to play the endgame with white (King and Rook) against the computer and checkmate the black King in as few moves as possible. The difficulty of the test exercises was comparable to the chess exercises in the learning phase, ranging from checkmate in 6 to checkmate in 10 moves. Participants were given a maximum of 15 minutes for each test situation. Following the rules of the World Chess Federation (FIDE), a game ended in a draw when 50 moves were made while no piece was captured, the black King took the white Rook (no checkmate possible), or the same position reoccurred three times in the same game.

Analysis

Chess principles

The predictions of the participants in the PSE and PO condition in the learning phase were compared to the computer's move. When a predicted move was identical to the move of the computer it was scored one, whereas predicted moves that were different were scored zero. Since there were sometimes several correct moves, the predicted move was also scored one if it was equally good as the computer move.

To acquire more specific information about the development of understanding of the principles, the predictions in the learning phase and the moves in the test exercises were analyzed blindly by a chess master (a 21-year old female member of the Dutch youth team). The chess master, together with the first author, identified the different principles that apply to the KRK endgame. Six principles were identified, such as limiting the space of the black King with the Rook, and minimizing the difference between the two Kings. In Figure 1, the principles are stated and explained. Although principle 1 and 2 look similar at first sight, principle 1 (i.e., the Rook limits the space of the King by driving it to the edge of the board) is only applicable at the start of the endgame. It is applied to minimize the space of the black King in general, by moving the Rook closer to the King. Principle 2 (White King takes opposition, Rook checks the black King) is a specific type of move used to directly force the black King to the edge of the board. This principle is applied towards the end of the endgame. In no cases do these two principles overlap.

First, the chess master analyzed for each time white had to move which of the six principles had to be applied and which move was best. Next, the chess master identified whether the actual prediction (in the learning phase) or move (in the test phase) of the participants indicated that they applied that specific principle. The total number of correctly applied principles was summed. A second chess master re-analyzed 40% of the data. Interrater agreeability exceeded .95. In cases of disagreement, raters jointly decided on the correct coding. In Figure 1, the principles are listed and demonstrated in an example of a chess position. Furthermore, the correct move for white is given, accompanied by how participants' responses were coded.

Figure 1

The chess principles that correspond with the King-Rook versus King endgame and corresponding examples of chess positions. The chess principles are formulated in the second column, the explanation of the principles is stated between brackets. When participants applied the correct principle this was coded as 1. When participants failed to apply the correct principle, this was coded as 0 (see column 5). Afterwards, the total number of moves that indicated correct use of principles was calculated. Predictions were only provided by the PSE and PO conditions. Note: The numbering of the principles does not imply a hierarchical structure.

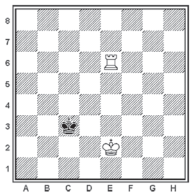
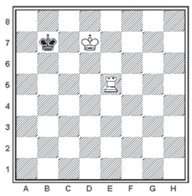
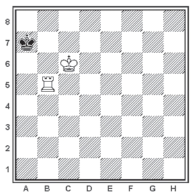
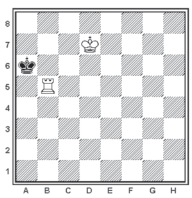
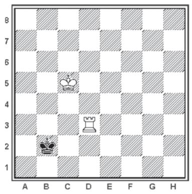
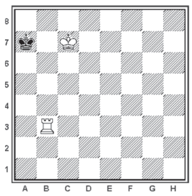
Position	*Principle*	*Correct move*	*Participant example response*	*Coding*
	1. The Rook limits the space of the King by driving it towards the edge of the board. [The Rook approaches the black King to minimize its space]	Rook e4	Rook g4	0
	2. White King takes opposition, Rook checks the black King. [The white King directly opposes the black King, while the Rook forces the black King to move towards the edge of the board]	Rook b5	Rook b5	1
	3. Tempo move by the Rook. [The rook moves up and down column B to prevent the black King from moving away from the edge of the board]	Rook either b1/b2/b3/b4	King c5	0
	4. Covering the Rook with the King. [To prevent the Rook from being captured by the black King, the white King adjoins the Rook. In this situation, the black King cannot capture the Rook, because he would be checked]	King c6	King c6	1

Figure 1 *(Continued)*

Position	*Principle*	*Correct move*	*Participant example response*	*Coding*
	5. Limiting the space between the two Kings. [The black King can only be checkmated if there is only one field separating the white and black King. Therefore, the white King moves up to the black king]	King c4	Rook d4	0
	6. Checkmating the King. [The white King is close enough to the black King to prevent it from moving away from the edge of the board. At that point, the Rook checks the black King, so he can no longer move at all without putting himself in check]	Rook a3	Rook a3	1

Self-explanation protocols

The coding of the think-aloud protocols was focused at identifying accurate self-explanations. Given the above-mentioned distinction between chess rules and chess principles, we examined the protocols for segments that corresponded with one of the following categories:

1. *Chess rule explanations*. The number of times that a participant referred to one of the basic chess rules, being either the legal moves of the King and Rook, capturing, check, checkmate and stalemate. For example: "I can't move the King from e1 to e3, because the King is only allowed to move one space." Or: "I have to move the Rook, because he is in check."
2. *Partial explanation of chess principles*. This category was scored if a participant referred to one of the chess principles, but only incompletely. For example, when someone stated: "The Rook goes to B7, so that the black King has less space." This actually refers to principle 1 in Figure 1 (The Rook limits the space of the King by driving it to the edge of the board), but only partially.
3. *Complete explanation of chess principles*. If a participant fully named one of the principles (see Figure 1), category 3 was scored. We hypothesized that this category would occur less frequently than category 2, since

participants who have only partial knowledge will rarely state complete explanations of the chess principles.

In sum, the development of principled understanding was analyzed at five points by examining a) the use of principles as indicated by the move predictions in the learning phase, b) the content of the self-explanations for the PSE condition (i.e., to what extent did the learners explicitly infer any of the principles?), c) the use of principles as indicated by the moves in the test exercises, d) the extent to which learners were able to transfer the acquired principles to a novel situation by checkmating the black King in the test exercises, and e) the number of procedural errors committed in the test exercises.

Results

Learning phase

To determine the study times per move for the PO and Observation condition, the study times per move for the PSE condition were calculated. Participants in the PSE condition needed on average 42.5 seconds (*SD* = 19.6) to predict the next move. After the computer generated the optimal move, participants needed on average 12.3 seconds (*SD* = 4.2) to evaluate their prediction and explain possible discrepancies between their prediction and the computer's move. The study times differed somewhat between the various chess exercises, with more complex situations resulting in longer study times. For analyses of variance, all alpha levels were set at .05. When sphericity assumptions were violated, this is mentioned in the text.

Number of correct move predictions

The differences between the PSE and PO condition in the number of correct predictions made in the learning phase were analyzed in a 2 x 2 repeated measures ANOVA, with study session (first or second) as the within-subjects factor, and condition (PSE or PO) as the between-subjects factor. In the learning phase, the Observation condition did not provide performance data. There was no effect of session, $F < 1$. However, the difference between the PSE and PO conditions was significant, $F(1, 28) = 6.71$, $MSE = 126.81$, $p < .05$, $\eta_p^2 = .19$. Participants in the PSE condition produced significantly more correct predictions than participants in the PO condition. The interaction effect was not significant, $F(1, 28) = 1.55$.

Principles applied in predictions

Apart from comparing the number of correct predictions, we also analyzed the content of the predictions to examine the development of understanding of the chess principles. The results showed that there was an effect of session, F (1, 28) = 7.22, MSE = 59.03, $p < .05$, $\eta_p^2 = .21$. In the second session, participants in both conditions applied the principles more than in the first session. Moreover, participants in the PSE condition applied the principles more than participants in the PO condition, F (1, 28) = 14.47, MSE = 164.35, $p < .001$, $\eta_p^2 = .34$. No interaction effect was found, $F < 1$. Table 1 presents the percentage of correct predictions per study session and the percentage of predictions that reflected the correct application of the principles.

Table 1

Mean percentage of correct move predictions and move predictions that expressed correct application of principles in chess exercises in the learning phase (standard deviations between brackets)

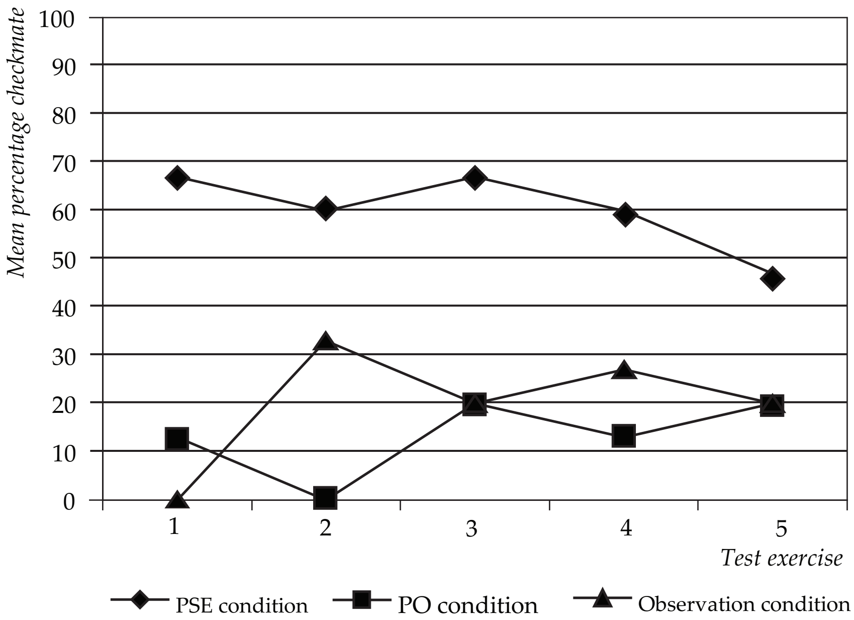
Session	PSE condition	PO condition
	Mean % of correct predictions	
1	64.8 (10.9)	59.3 (9.9)
2	67.5 (6.5)	57.9 (8.6)
Total Mean	66.1 (8.7)	58.6 (9.2)
	Mean % of correct principles	
1	58.9 (10.8)	48.4 (9.7)
2	66.3 (10.1)	51.1 (11.6)
Total Mean	62.6 (10.5)	49.7 (10.6)

Test phase

Number of checkmates

A 5 x 3 repeated measures ANOVA with test exercise (1, 2, 3, 4, 5) as the within-subjects factor, and condition (PSE, PO, Observation) as the between-subjects factor revealed that, although there was no effect of test exercise, $F < 1$, the three conditions differed significantly in number of checkmates in the test exercises, F (2, 42) = 11.53, MSE = 4.49, $p < .001$, $\eta_p^2 = .35$. Post hoc tests using Bonferroni's method indicated that the PSE condition checkmated the King significantly more often than the PO condition, $p < .001$, and the Observation condition, $p < .01$. The difference between the PO and Observation condition was not significant, $p > .05$. No interaction effect was found, $F < 1$. The mean number of checkmates for the PSE condition was 3.0 (*SD* = 1.77), whereas for the PO and Observation condition this was 0.87 (*SD* = 0.92) and 1.33 (*SD* = 1.68), respectively. In Figure 2, the percentages of checkmate per test exercise are represented per condition.

Figure 2
Mean percentage of participants who checkmated the black King in the test phase per condition



Principles used in test exercises

A repeated measures ANOVA revealed that although there was no effect of test exercise, $F < 1$, a significant difference in the correct use of the principles relevant to the KRK endgame was found, F (2, 42) = 5.02, *MSE* = 4603.46, $p < .05$, $\eta_p^2 = .19$. Post hoc analyses indicated that participants in the PSE condition more often applied the principles correctly than participants in the PO condition and the observation condition. No interaction effect was found, $F < 1$.

Table 2

Mean percentage of correct use of chess principles in the test exercises (standard deviations between brackets)

Test exercise	PSE condition	PO condition	Observation condition
1	62.39 (25.46)	49.85 (21.69)	46.14 (27.27)
2	62.91 (20.69)	50.35 (15.38)	52.99 (18.28)
3	67.29 (14.61)	48.75 (15.80)	49.32 (20.38)
4	63.04 (19.31)	51.48 (15.23)	48.54 (14.96)
5	55.43 (16.67)	40.05 (19.71)	49.33 (15.21)
Total mean	62.21 (14.24)	48.10 (9.35)	49.26 (16.11)

Errors in test exercises

When participants were able to checkmate the black King, no differences were found between conditions in the number of moves needed to checkmate, $F(2, 26) = 1.16$.

Finally, the number of 'procedural errors' in the test exercises was also identified. This category covered the number of times a game ended because 50 moves were made without checkmate or capturing a piece, 15 minutes had passed, the Rook had been captured, or the same position had occurred three times. A one-way ANOVA with condition as the independent variable and number of procedural errors as the dependent variable indicated that there was a significant difference in the number of procedural errors between conditions, $F(2, 42) = 3.86$, $MSE = 6.96$, $p < .05$, $\eta_p^2 = .16$. Post hoc analysis revealed that the PSE condition made fewer procedural errors than the PO condition. Participants in the PSE condition committed on average 1.5 procedural errors ($SD = 1.3$), while for the PO and Observation condition this was 2.8 ($SD = 1.2$) and 2.6 ($SD = 1.5$), respectively.

Although the conditions differed in the number of times checkmate and in the correct use of principles, there were no differences in the time participants needed to select their next move in the test exercises, $F < 1$. Participants in the PSE, PO and Observation condition needed on average respectively 22.9 seconds ($SD = 7.7$), 23.4 seconds ($SD = 8.0$), and 21.0 seconds ($SD = 7.6$) to select their next move. This indicates that although participants in the PO and Observation condition performed worse than the PSE condition in the test phase, this was probably not due to a lack of motivation to study the test exercises thoroughly (Bruinsma, 2003).

Self-explanations in the learning phase

The test phase results indicated that the PSE instruction had a more positive learning effect than the PO instruction. To gather more information on how the PSE instruction affected learning, we analyzed the self-explanation protocols of the participants in the PSE condition. Over the 10 chess exercises, participants explicated on average 99.7 (*SD* = 56.3) self-explanations. Of these self-explanations 76.4 (*SD* = 43.9) were verbalized when predicting the next move, while 23.3 (*SD* = 12.7) were generated when explaining the discrepancy between the participant's move and the computer move (i.e., the evaluation of computer feedback). When correlating the number of self-explanations with performance in the test phase (i.e., the number of times checkmate in the test phase), we found that particularly the self-explanations made when predicting the next move were related to performance in the test phase ($r = .56$, $p < .05$). The correlation between number of self-explanations during the computer feedback evaluation and number of times checkmate in the test was .46 ($p = .09$). Therefore, we decided to perform a median split on number of self-explanations made when predicting the next move. The high-explainers (N = 7) consisted of all participants who provided more than 51 self-explanations (median was 51) when predicting the next move. Their mean number of self-explanations was 95.1 (*SD* = 43.5). The low-explainers group (N = 8, mean number of self-explanations = 32.4, *SD* = 11.5) all generated less than 51 self-explanations when predicting the next move. As to the time participants needed to self-explain the next move, the difference between the high- and low-explainers was not significant, F (1, 13) = 3.69, $MSE = 436.61$, $p = .08$, $\eta_p^2 = .22$. The same was found for the time needed per move to explain the discrepancy between prediction and computer move during the computer feedback evaluation, $F < 1$.

Principles named in self-explanations

To examine a possible development in the content of self-explanations, we summed the self-explanations in the first session (exercise 1-4) and the second session (exercise 5-10) and ran separate 2 x 2 repeated measures ANOVAs on all three self-explanation categories, with session (first or second) as the within-subjects factor, and self-explanation group (low or high) as the between-subjects factor. Since participants in the PSE group had no time restriction when self-explaining, we entered total time used for self-explaining as a covariate in the analyses to control for the effect of differential time on task. The covariate effect will only be mentioned when significant. For category 1 explanations (explanations referring to chess rules), no effect of session was found, $F < 1$. However, the group effect was significant, in favor of the high-explainers, F (1, 12) = 5.07, $MSE = 478.64$, $p < .05$, $\eta_p^2 = .30$. The interaction effect was not significant, $F < 1$. For category 2 partial explanations, the effect of session was nearly significant: In the second session more category 2 self-explanations were generated than in the first session, F (1, 12) = 4.63, MSE = 92.54, p = .053, $\eta_p^2 = .28$. The group effect was also significant: the high-explainers provided more partial explanations of the chess principles than the low-explainers, F (1, 12) = 8.40, MSE = 137.48, $p < .05$, $\eta_p^2 = .41$. Finally, for category 2 self-explanations, the interaction between session and group was significant, F (1, 12) = 5.09, MSE = 92.54, $p < .05$, $\eta_p^2 = .30$. This interaction effect indicated that the difference between the high- and low-explainers in the number of partial explanations of the chess principles that was already significant in session 1 increased in the second session. The partial explanations of the chess principles (category 2), although they do not consist of perfect verbalizations of the chess principles, describe the essence of the chess principle and indicate advanced understanding of the principles. For the complete explanations of chess principles (category 3 explanations), no effect of session was found, $F < 1$. However, the difference between the high- and low-explainers in category 3 explanations was significant, F (1, 12) = 6.15, MSE = 8.93, $p < .05$, $\eta_p^2 = .34$. The interaction effect was not significant, $F < 1$. In Table 3 the number of self-explanations per category and per session are shown for both the high- and low-explainers.

Table 3

Mean number of self-explanations, divided over three categories of self-explanations, for the high- and low-explainers (standard deviations between brackets)

Session	High-explainers	Low-explainers
	1. Chess rule explanations	
1	32.29 (20.38)	8.63 (7.25)
2	31.43 (26.50)	7.25 (3.54)
Total	64.71 (45.36)	15.75 (6.73)
	2. Partial explanation of chess principles	
1	12.71 (6.02)	8.75 (4.74)
2	34.71 (17.63)	16.25 (10.81)
Total	47.42 (20.31)	25.00 (13.04)
	3. Complete explanation of chess principles	
1	5.71 (3.15)	3.13 (2.23)
2	6.43 (4.04)	5.38 (2.62)
Total	12.14 (5.00)	8.50 (3.34)
Grand Total	124.27 (53.97)	49.25 (14.99)

Self-explanations and test phase

To examine whether high- and low-explainers differed in performance on the test exercises, a 5 x 2 repeated measures ANOVA was performed on the data, with test exercise as the within-subjects factor and self-explanation group (high or low) as the between-subjects factor. Although the test exercise effect was not significant, $F < 1$, the between-subjects difference was: the high-explainers checkmated the King more often than the low-explainers, F (1, 13) = 6.54, MSE = 143.60, $p < .05$, $\eta_p^2 = .34$. The interaction effect was not significant, $F < 1$. We ran a one-way ANOVA on procedural error data, with high- versus low-explainers as the between-subjects factor and number of procedural errors committed in the test phase as the dependent variable. The high-explainers committed significantly less procedural errors than the low-explainers, F (1, 13) = 5.53, MSE = 2.38, $p < .05$, $\eta_p^2 = .30$. The high-explainers made on average 1.00 procedural errors (SD = 1.16), whereas the low-explainers committed on average 2.88 errors (SD = 1.81) in the test exercises.

Discussion

The results of this study suggest that predicting the next move combined with self-explaining the predictions positively contributed to the development of principled understanding of a chess endgame. Participants in the prediction and self-explanation condition (PSE condition) showed better understanding of the principles that underlie the KRK endgame than the prediction only condition (PO condition). The higher level of principled understanding was exemplified at five moments. First, the PSE condition more often predicted moves that exemplified correct application of the principles in the learning phase. Furthermore, the PSE condition was in more test exercises able to apply the inferred principles to novel problems and checkmate the black King than the PO or Observation conditions. Moreover, the test results indicated that the PSE group committed less procedural errors. In sum, the test performance of the PSE condition indicates that participants had developed principled understanding that allowed them to transfer the inferred principles to a novel task (Gelman & Greeno, 1989). These results further strengthen the evidence for the self-explanation effect (Chi et al., 1989; Chi et al., 1994; Renkl, 1997). This study showed that asking learners to self-explain their move predictions positively affected development of principled understanding.

The analysis of the content of the self-explanations provides further insight into how principled understanding emerged. Since participants had no knowledge of the principles of the endgame, but only learned the basic chess rules at the start of the experiment, self-explanations were mostly focused at the explication of basic chess rules (e.g., the legal moves of the Rook and King, naming of situations of check). In both study sessions, the high-explainers in the PSE condition explicated more references to basic chess rules than the low-explainers. However, already in the first session, the high-explainers made more partial explanations of the chess principles. In the second session this difference increased as indicated by the interaction effect. Whether the higher explication of basic chess rules caused the increase in naming of partial principle explanations in the high-explainers is at this point unclear. In future research, rehearsal of basic chess rules should be manipulated to provide insight into this issue.

Previous research showed that the self-explanation effect is mostly found when using participants of at least an intermediate phase of skill acquisition. In the present study, novices without any chess knowledge were able to profit from the prediction and self-explanation instruction. A possible explanation for this discrepancy is found in the nature of the study material. When learners have to generate self-explanations of study material that is to a certain extent of a textual nature, it is possible that the self-explanations remain at a reread or problem restatement level, without meaningfully explaining the study material. For instance, Mwangi and Sweller (1998) found that about 39% of the self-explanations in their study could be classified as rereads. In the present study, extensive restating of the problem situation was unlikely, because no explicit description of the problem situation or of the solution procedure was provided. The only information participants received were a chess position and the requirement to predict and explain the next move of the computer. Therefore, instead of remaining at a reread level, participants were stimulated to provide self-explanation inferences that go beyond restatement of information and indicate inference of new knowledge (Chi, 2000). This, in turn, enhanced discovery of the endgame principles. In the current study, category 2 and 3 explanations can be considered self-explanation inferences, since they involved (partial) explanations of the chess principles that were nowhere stated, but had to be inferred by the learners. According to this guideline, for the high-explainers approximately 70% of all self-explanations can be classified as self-explanation inferences, whereas for the low-explainers this was 75%. The absence of an explicit problem description might in part explain the high percentage of self-explanation inferences.

Moreover, the wording of the self-explanation instruction in this study might have played a role. Participants in the PSE condition were asked to explain *why* they thought the computer would make the move they predicted. The *why* question asked for a meaningful, reason-based explanation of the prediction (see also Calin-Jageman & Horn Ratner, 2005; Chi et al., 1994), and decreased the likelihood that participants merely restated the position of the pieces on the chessboard. This interpretation is supported by two findings. First, the PO instruction that did not require learners to provide why-type explanations for their predictions had a marginal learning effect compared to the PSE instruction. Second, learners who provided more self-explanations in the PSE condition developed better principled understanding as indicated by results from both the learning and test phase. Possibly, the wording of the self-explanation instruction in this study caused learners to reflect on the reasons for their predictions, which led to better development of principled understanding. To test this assumption, more research is needed that compares the wording of the self-explanation instruction used in this study to other formulations of self-explanation instructions.

It is possible that participants in the PO condition to a certain extent covertly self-explained their predictions, because the instruction to predict the next move asks for application of principle knowledge. Despite the possible covert self-explanation, the development of principled understanding of the PO group lagged behind the PSE group. One explanation is that the PO instruction did not promote active involvement in the learning task. However, the results on move predictions and application of principles in the learning phase indicate that the PO group shows a learning curve in the development of principled understanding. Therefore, the performance difference between the PSE and PO condition supports the conclusion that verbalization of self-explanations out loud is necessary to trigger the full scope of the self-explanation effect. A future study in which the covert self-explanation activities of the PO group are monitored can provide information on the differential effect of overt and covert self-explanation.

It would also be interesting to investigate in future research how the computer feedback stimulated the learning process. For example, Renkl (1999, 2002) mentions that having students generate self-explanations without providing feedback can lead to 'illusions of understanding.' If students do not receive feedback on the accuracy of the content of their self-explanations, they are not encouraged to identify and correct misconceptions and might persevere in illusions of understanding. In the present study, we again found that although self-explaining enhanced the development of principled understanding, learners were not always able to name the complete, or in some cases even the partial principle. Moreover, especially the low-explainers committed a considerable amount of procedural errors in the test phase. Providing learners with feedback about the quality of the content of their self-explanations might be a possibility to immediately correct misconceptions and prevent illusions of understanding. Also, extending the use of self-explanations with the use of instructional explanations (i.e., explanations generated by more knowledgeable persons such as teachers) that are appropriately timed could aid in reducing the occurrence of illusions of understanding (Renkl, 2002).

Although this is a first study on the self-explanation effect in novices, and future research is needed to gain insight into the cognitive mechanisms that explain our findings, implications for education emerge. For example, verbalizing possible reasons for solution steps when studying a task, even when it is not of a textual nature, is an instructional procedure that is easy to implement and can evoke better understanding of the material. Our findings discourage instructing learners to predict next solution steps, when this is not combined with an instruction to verbally explain the predictions. Finally, it is important to stimulate learners to focus on reason-based explanations for their predictions, instead of only restating the problem situation. However, before proceeding to the implementation of these instructional techniques, research is needed that tests the scope of the self-explanation effect in novices in other fields besides chess.

Chapter 5

Monitoring accuracy and self-regulation when learning to play a chess endgame[4]

[4] This chapter was published as: De Bruin, A.B.H., Rikers, R.M.J.P, & Schmidt, H.G. (2005a). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology, 19*, 167-181.

Abstract

In order to examine the effect of monitoring and self-regulation on skill acquisition, the present study asked novice chess players to provide judgments of learning (JOLs) and to select moves for restudy after studying an endgame of chess. In four groups, we varied the JOL instruction (present versus absent) and the selection instruction (free number of move selections versus selection of at least two moves per chess exercise). After four learning trials, participants were required to play against a chess computer. In the learning phase, participants who were forced to select moves for restudy outperformed those who were free to select moves for restudy when predicting the next computer move, even after controlling for actual number of restudied moves. Although the groups that did provide JOLs showed better self-regulatory behavior, there were no or even negative performance differences between the groups that did provide JOLs and the groups that did not provide JOLs. This same pattern emerged in the test phase: Although no differences were found between the groups with and without JOLs, the groups that were forced to select moves for restudy outperformed the groups that were free in number of move selections. These data show that, for novice chess players, the instruction to provide JOLs possibly places a high and ineffective load on working memory and therefore has no effect on learning a chess endgame. To examine the relation between prior knowledge and quality of self-regulation, further research is needed that examines the effect of the JOL and selection instruction in groups that differ in chess experience.

In recent years, the interest in metacognition in cognitive research has grown considerably (for reviews, see Hacker, Dunlosky, & Graesser, 1998; Metcalfe & Shimamura, 1994; Schunk & Zimmerman, 1994). Metacognition is defined as awareness and control of one's learning (Baker & Brown, 1984). Metacognitive processes are the control processes that direct more basic learning processes and include knowing how to use available information to reach a goal, judging the cognitive demands of a particular task, and assessing one's progress during and after performance (Flavell, 1987). Metacognition is typically associated with experts, who make use of elaborate metacognitive strategies, such as self-monitoring (Do I understand what I am studying?) and self-regulation (What part of the material needs restudying?) when solving a problem (Vu, Hanley, Strybel, & Proctor, 2000). Consider musicians who are struggling with a highly complex music piece. Improvement in play depends on their capacity to analyze crucial parts in the music piece and identify and correct their own weaknesses. Only by deliberately reflecting on the practice session they will be able to further develop the quality of their performance.

Research on metacognitive activities in experts has shown that expert computer programmers show more metacognitive awareness of their performance than novices, and possess better self-regulating strategies such as monitoring their progress in learning and regulating performance (Eteläpelto, 1993; Shaft, 1995). Moreover, Schönfeld (1987) found that expert mathematicians needed more time to analyze a problem situation, and were constantly monitoring their comprehension level, whereas novices showed little use of spontaneous self-monitoring and self-regulation. Long and Long (1987) found that successful readers distinguished themselves from less successful readers in the extent to which they actively tested their comprehension level, by anticipating questions and summarizing information during reading.

Although metacognitive strategies are complex and increase working memory load, several studies have shown that they can be successfully taught to inexperienced learners (e.g., Azevedo & Cromlye, 2004; Guthrie, Wigfield, & VonSecker, 2000). A study by Leonesio and Nelson (1990) on learning verbal material indicated that participants' judgments of knowing after having studied a list of words positively correlated with subsequent recall. Moreover, Schraw (1994) found that students make spontaneous use of metacognitive skills with varying success: Students who were high self-monitors were more confident about their performance, and performed better than so-called low self-monitors.

As to experimental research on the effect of regulation of study on later test performance, Thiede (1999; Thiede & Dunlosky, 1999) asked participants to make a judgment of learning (JOL) after studying a list of 36 Swahili-

English translation equivalents. That is, participants were asked to make a prospective judgment of the likelihood of correctly recalling an item on the list. A measure of monitoring accuracy was calculated by correlating JOLs and actual recall: A positive correlation indicates better monitoring accuracy. Furthermore, self-regulation of study was examined by computing the correlation between JOLs and selection of items for restudy. A negative correlation indicated accurate self-regulation, since participants selected those items for restudy that they considered difficult. Their findings indicated that both monitoring accuracy and self-regulation were reliably related to test performance, although the number of items selected for restudy was not. Apparently, test performance was more affected by the extent to which students selected items that needed restudy than by how many items students restudied. In general, their findings demonstrate that instructing students to discriminate between what they know and what they do not know can be a successful metacognitive strategy that improves learning through its effect on self-regulation of study.

Until now, the judgment of learning paradigm has been applied to learning lists of words and, in an adapted version, learning from expository texts (Thiede, 1999; Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003; Thiede & Dunlosky, 1999). Since the JOL paradigm has a straightforward design that requires little instruction, it is interesting to further examine the effects of this metacognitive procedure on learning of more complex material. In the present study, we applied the JOL paradigm as used by Thiede and Dunlosky (1999; Thiede & Dunlosky, 1999) in cognitive skill acquisition, namely learning to play a chess endgame. By asking participants to judge their understanding of the studied chess moves and providing the opportunity for restudy of moves, the JOL paradigm can be applied to a complex task as playing chess.

Four different groups (all novices in chess) were tested to examine the effects of providing JOLs and selecting material for restudy on later test performance. After studying a computer presentation about the basic rules of chess, participants were shown a computer chess endgame of Queen and King (white pieces) versus King (black piece), and were required to predict the move the computer would make every time white had to move. After the game had ended, participants in the first group were required to judge, for every time white had to move, how confident they were that they would accurately predict the correct move. To monitor the accuracy of their judgments, participants were afterwards asked to predict the moves of a mirror image of the same chess exercise. Finally, participants were given the opportunity to select moves for restudy. By varying the JOL instruction (present versus absent) and the move selection instruction (free number of move selections versus selection of at least two moves per chess exercise), we were able to test the effects of the self-monitoring and self-regulating activities. Participants worked through four learning trials, consisting of four King-Queen versus King (KQK) endgames. After the learning phase had finished, participants had to play four new versions of the endgame against the computer, and checkmate the opponent king in as few moves as possible.

Given the absence of prior knowledge, and hence the difficulty for learners to self-regulate their learning (i.e., select moves for restudy), we hypothesized that the two groups that were forced to select moves would be encouraged to more extensive self-regulation by selecting moves for restudy that were not fully understood as indicated by their judgments of learning. This would probably have a positive effect on performance in both the learning phase and the test. Moreover, providing JOLs stimulates metacognitive awareness and might interact with the forced move selection instruction, aiding learners in the selection of moves for restudy and hence in performance development. Overall, we hypothesized, on the one hand, a positive correlation between JOLs and predictions, and, on the other hand, a negative correlation between JOLs and move selections for restudy (Thiede, 1999).

Method

Participants

Participants were 73 undergraduate students from Erasmus University Rotterdam, The Netherlands (*M* age = 20.2, *SD* = 2.1). Only students who had never played chess before, nor were aware of the rules of chess, were allowed to participate. Participants were randomly assigned to one of four groups (18 in every group, except for the No JOL free selection group, which contained 19 participants). All students participated to fulfill course requirements, and were awarded a small financial compensation.

Materials

At the start of the experiment, participants studied a computerized presentation about the basic rules of chess. On a computer screen the rules regarding the legal moves for King and Queen were visualized, as well as the basic rules of capturing, check, checkmate, and stalemate. It is important to note that these rules can be clearly distinguished from the principles that underlie the endgame: Knowing what the legal moves of a Queen and King are does not tell a learner anything about how to effectively play this endgame. Afterwards, participants completed a short paper-and-pencil test, consisting of nine true-false questions, to assess whether they had understood the rules.

The chess computer program ‘Fritz 7’ provided the chess exercises for the learning phase and the test phase, which consisted of examples of the endgame of King and Queen against King. This chess program plays at Grandmaster level (ELO rating of approximately 2700). The difficulty level of the chess exercises was held constant throughout the experiment: All exercises were examples of endgames of checkmate in six moves, consisting of the same underlying chess principles. These chess exercises were presented on a computer screen, embedded in the context of a computer application. The application recorded the participants’ move predictions, JOLs, and move selections. In the test phase, participants played four versions of the endgame against the computer program Fritz 7. The test exercises ranged in difficulty from checkmate in six to checkmate in nine moves.

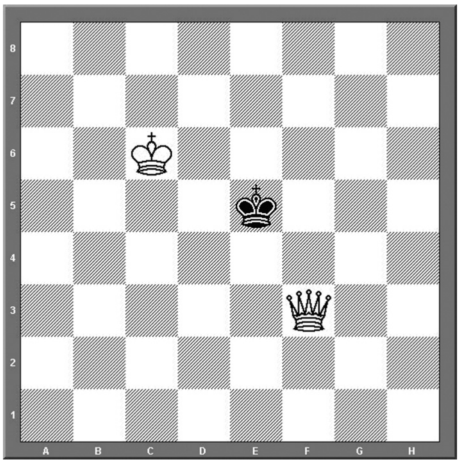
Procedure

Participants started by studying the computer presentation about the basic rules of chess in a self-paced way. This took on average about 25 minutes. Afterwards, they completed a paper and pencil test to assess their degree of understanding. Any uncertainties were explained immediately by the experimenter before proceeding to the learning phase. No differences existed between groups on this test, F (3, 69) = 1.15, MSE = 0.79, $p > .05$.

Since it was impossible to learn the chess game completely during the length of one experiment, a subpart of the game was selected that could be studied in isolation. This subpart involved the theoretical endgame of Queen and King against King (KQK endgame). A theoretical endgame is an endgame that has a fixed result: The winner in the KQK endgame is the white side with King and Queen. The task of the white side is to correctly play and win the endgame according to the particular endgame’s principles. The chess exercises of the learning phase were designed to enable participants to infer the principles that are essential to the KQK endgame. These principles were nowhere stated, but had to be discovered by studying the computer moves in the chess exercises. Participants worked through the application individually, which lasted on average an hour. An example of a chess position used in the experiment is provided in Figure 1.

Figure 1

Example of a chess position used in the learning phase



We begin by describing the basic outline of the experiment. All groups started by studying a chess exercise that required them to predict the moves that the white side (Queen and King in this study) was going to make. Previous research (Chi, De Leeuw, Chiu, & LaVancher, 1994; De Bruin, Rikers, & Schmidt, in press; Renkl, 1997) has shown that generating self-explanations during self-study has a positive effect on learning. To promote thorough processing, we instructed participants in this phase of the experiment to provide self-explanations while predicting the next move. Participants received feedback on the accuracy of their predictions by observing the actual moves of the computer. Based on a similar manipulation in a previous study (De Bruin et al., in press), participants were given 40 seconds for every move prediction. When participants had predicted all six moves of the particular endgame, they proceeded to the Judgment of Learning phase. In this phase, participants were shown all positions from the previously studied chess exercise at which white had to move. To prevent any learning of the sequence of moves in this phase, they were presented in random order. For every move, participants were asked to judge their learning by answering the following question and pressing a corresponding button: "How confident are you that you will predict the correct move?" (0% = definitely will not be able to predict, 20%, 40%, 60%, 80%, and 100% = definitely will be able to predict). No self-explanations were generated during this phase. After having provided comprehension ratings for all moves, participants were shown a mirror image of the same chess exercise and were once more asked to predict the next moves for white (i.e., the transfer phase) so they could monitor the accuracy of their confidence ratings. A mirror image exercise used the same starting position and subsequent moves as the exercise in the prediction phase, except that the chess pieces were arranged in a mirror image form compared to the prediction phase. We used a mirror image of the exercise to avoid participants from recalling the correct moves based on memorization instead of correct application of the principles. Since this phase of the experiment was intended as a monitoring accuracy phase and not as a learning phase, participants were not asked to self-explain and were given only 20 seconds per move to predict the computer moves.

Finally, participants were allowed to select and restudy the moves of the white side of the mirror image exercise. Participants were shown all positions of the mirror image exercise at which white was to move one by one, and were asked to press 'y' if they wanted to select the particular move for restudy, and 'n' if not. No self-explanations were provided during this phase. The restudy phase comprised studying those moves that were selected in the selection phase, but now of another mirror image of the same exercise, under the prediction and self-explanation instructions equal to the first part of the chess exercise. If a participant selected move one, three, and five, only these moves were studied in the restudy phase, in the order in which they appeared in the chess exercise. After participants had worked through all of these phases, a new chess exercise of the same difficulty level was presented and they were once again asked to predict the moves of the computer, afterwards provide JOLs, followed by a transfer phase, etc. Participants studied a total of four chess exercises under these instructions.

To stimulate discovery of the chess principles particular to the KQK endgame, the basic outline of the experiment was altered to manipulate self-monitoring and self-regulation. The structure of the learning phase of the experiment was modified to create four conditions in a 2 x 2 design, varying the JOL instruction and the move selection instruction. In the first condition (i.e., the JOL free selection condition) participants received the instruction to provide JOLs, and were free as to the number of moves selected for restudy. This latter term is hereafter referred to as 'free move selection'. To examine the possible differential effect of forced versus free self-regulation on performance, participants in the second condition (i.e., the JOL forced selection condition) were forced to select at least two of the six moves for every chess exercise. This latter term is hereafter referred to as 'forced move selection'. To assess the influence of providing JOLs on self-regulation and performance, participants in the third condition did not rate their understanding by providing JOLs, but were only instructed to select at least two moves per chess exercise for restudy (i.e., the No JOL forced selection condition). Finally, a fourth group did not provide JOLs and was free in the number of move selections for restudy (i.e., the No JOL free selection condition). In sum, of the four groups, two provided JOLs and two did not, two groups were forced to select at least two moves for restudy, whereas the remaining two groups were free in the number of move selections.

After the four learning cycles of the experiment, participants were required to play four new KQK endgames with white against the computer and checkmate the black King in as few moves as they possibly could. Two of these test exercises were comparable in difficulty to the learning phase (checkmate in six moves), whereas the final two were slightly more difficult (checkmate in eight and nine moves). Participants received maximally five minutes to finish each test exercise. Following the rules of the World Chess Federation (FIDE), participants were instructed that a game ended in a draw when 50 moves were made and no piece was taken, the black King took the white Queen (checkmate impossible), stalemate occurred, or the same position occurred three times in the same game.

Analysis

The predictions of the participants in the prediction phase of each chess exercise were compared to the computer's optimal move. When a predicted move was identical to or as good as the move of the computer (sometimes several good moves were possible) two points were awarded, whereas predicted moves that led to a game situation that needed one more move to checkmate the black King were scored one. All other predictions (i.e., leading to situations that required two or more moves to checkmate the King) were scored zero. Maximally twelve points per prediction phase were awarded. The prediction scores are represented as percentages of the maximum score of twelve points. This scoring procedure was also applied to the predictions made in the transfer phase and the restudy phase of the experiment. Throughout the learning phase, participants received four prediction scores, four transfer scores, and maximally four restudy scores.

Repeated measures ANOVAs were applied to the data. First, in a 4 x 2 analysis the difference between the JOL free and JOL forced groups in level of JOLs was examined, with chess exercise (first, second, third, fourth) as the within-subjects factor and move selection (free or forced) as the between-subjects factor. Second, in a 4 x 2 x 2 analysis, with chess exercise (first, second, third, fourth) as the within-subjects factor, and JOL (present or absent) and move selection (free or forced) as the between-subjects factors, the differences in number of move selections, prediction scores, transfer scores, restudy scores, and test performance were analyzed. In these analyses, the differential effects of the JOL instruction and the move selection instruction were separately assessed. Specific details of the analyses are provided in the results section. To allow for comparison between conditions, post-hoc analyses were conducted using Bonferroni's method. Interaction effects will only be mentioned when significant. To assess the relation between the self-regulation measures (JOLs and move selections for restudy) and performance measures within groups, bivariate correlations were calculated between JOLs and move selections on one hand, and prediction scores, and transfer scores on the other hand. All alpha levels were set at .05.

Results

Learning phase

Judgments of Learning and move selections
To analyze differences in JOLs, a repeated measures analysis of variance with chess exercise (first, second, third, fourth) as the within-subjects factor and condition (JOL free selection, JOL forced selection) as the between-subjects factor was used. No effect of chess exercise, $F < 1$, but a significant interaction effect between group and chess exercise was found, F (3, 102) = 3.05, MSE = 104.56, $p < .05$. Further analysis showed that the interaction effect was of a quadratic form, F (1, 34) = 1,66, MSE = 120.85, $p < .05$, which means that the tendency of the differences between JOLs over time is characterized by one bend: the difference was small at the start of the learning phase, then increased, and finally lowered at the end (See Figure 2). No effect of group was found, $F < 1$. Mean JOL in the JOL free selection group was 45.9% (SD = 15.5), and 41.5% (SD = 12.2) in the JOL forced selection group.

In a repeated measures analysis of variance with chess exercise (first, second, third, fourth) as the within-subjects factor, and JOL (present or absent) and move selection (free or forced) as the between-subjects factors, the number of move selections for restudy showed a significant effect of chess exercise, F (3, 207) = 30.04, MSE = 1.98, $p < .001$, a significant interaction between chess exercise and JOL, F (3, 67) = 8.26, MSE = 16.31, $p < .001$, and a significant main effect of move selection, F (1, 69) = 34.73, MSE = 6.55, $p < .001$. Post-hoc tests using Bonferroni's method revealed that the two forced move selection groups (i.e., the JOL forced selection group and the No JOL forced selection group) selected significantly more moves than the two free move selection groups ($p < .001$). The interaction effect indicated that the two JOL groups started with a higher number of selections, but lowered these over the course of the exercises to the same level as the two No JOL groups. Mean number of selections were 1.08 (SD = 1.1) for the JOL free selection group, 2.08 (SD = 1.1) for the No JOL free selection group, 3.36 (SD = 1.4) for the JOL forced selection group, and 3.33 (SD = 1.5) for the No JOL forced selection group. In Figure 3 the move selection data are represented as a function of the between-subjects factor JOL (present or absent) and move selection (free or forced).

Figure 2

Mean Judgment of Learning per chess exercise. Percentages indicate how confident participants were that they would accurately predict a similar move

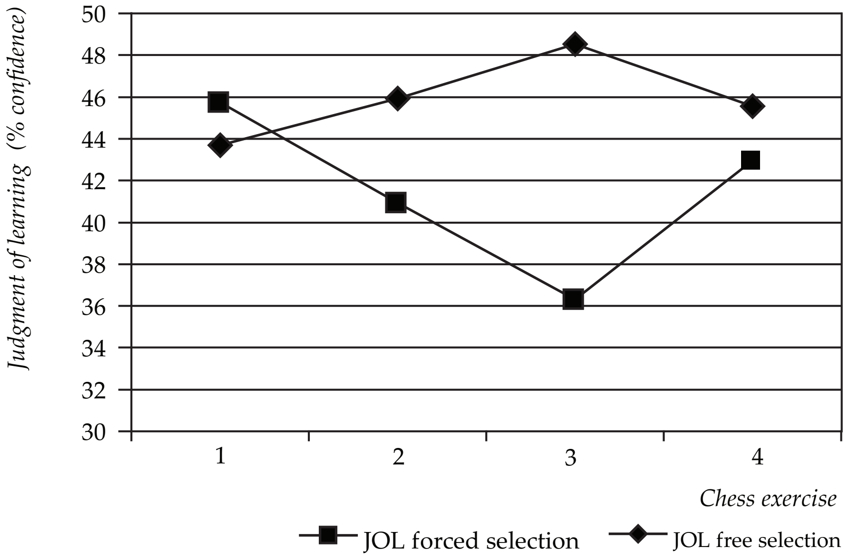
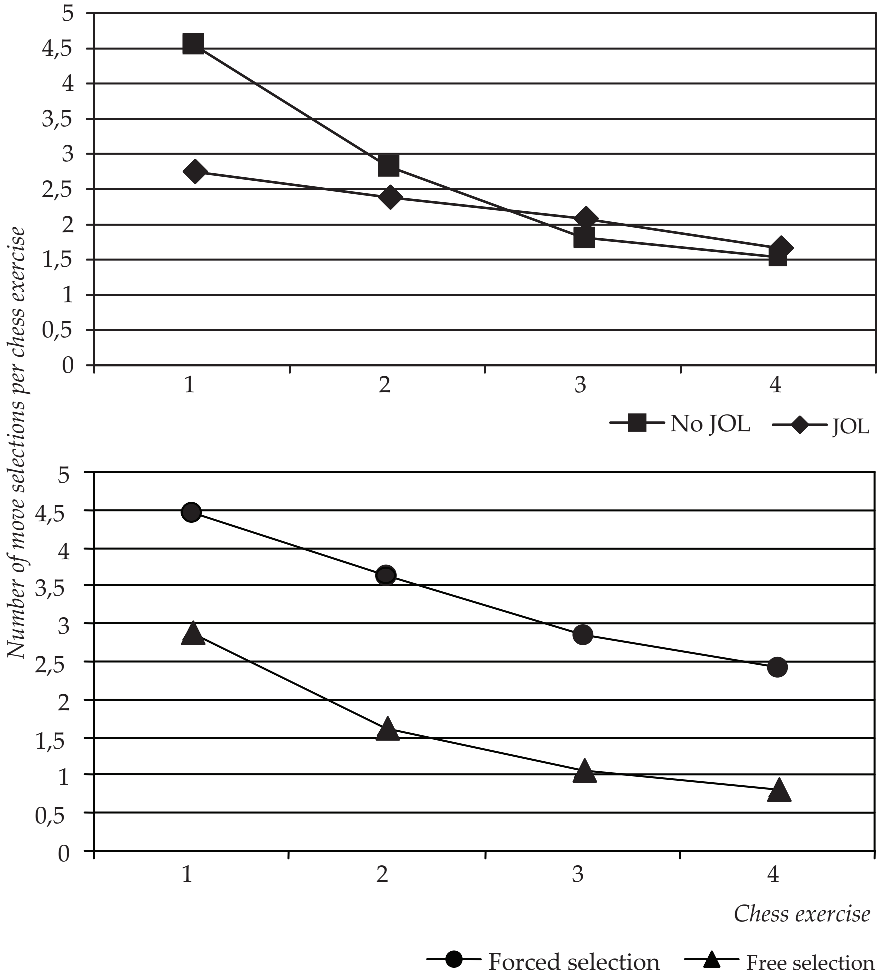


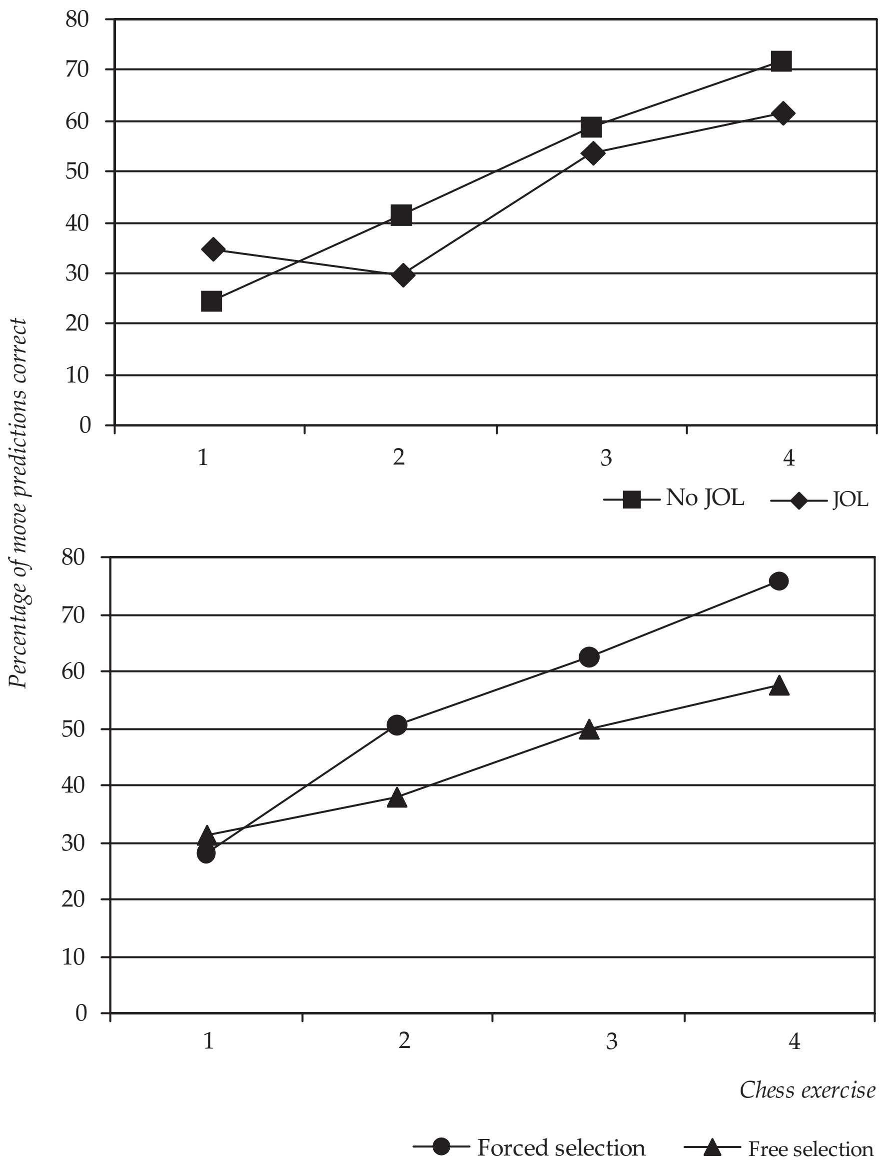
Figure 3
Mean number of move selections per chess exercise per group



Prediction scores in the learning phase
Since the selection data showed that the groups that were forced to select at least two moves for restudy actually had more practice because of their higher number of selections and hence number of restudies, it is logical to assume that these groups would perform better in the learning phase and the test phase than the two groups that were free in the number of move selections. To control for the effect of more extensive practice due to more move selections and hence more restudy, we incorporated the mean number of selections over the four chess exercises as a covariate in all analyses on the

performance data (prediction scores in prediction phase, transfer phase, and restudy phase). This analysis ensures that the differences in all prediction scores are corrected for the more extensive practice in the forced selection groups. Covariate effects will only be mentioned when significant.

Figure 4
Mean prediction score per chess exercise per group



In a repeated measures analysis of variance on the scores in the prediction phase with chess exercise (first, second, third, fourth) as the within-subjects factor, and JOL (present or absent) and move selection (free or forced) as the between-subjects factors, a significant effect of chess exercise was observed, F (3, 204) = 17.81, MSE = 230.53, $p < .001$, indicating that prediction scores improved over the course of the chess exercises. Moreover, a significant interaction effect between chess exercise and JOL, F (3, 204) = 6.61, MSE = 230.53, $p < .001$, and between chess exercise and move selection, F (3, 204) = 5,91, MSE = 230.53, $p < .01$, was observed. The interaction between chess exercise and JOL indicated that the No JOL groups increased more than the JOL groups. The interaction between chess exercise and move selection indicated that the forced move selection groups increased more than the free move selection groups. Finally, the overall difference between the groups was significant only for the move selection factor, F (1, 68) = 17.83, MSE = 405.12, $p < .001$. Post-hoc tests showed that the two free move selection groups (JOL free selection and No JOL free selection) had lower prediction scores than the two forced move selection groups, $p < .001$. An overview of all prediction scores in the learning phase is provided in Table 1 as a function of the between-subjects factor JOL (present or absent) and move selection (free or forced).

The same analysis was run on the move predictions made in the transfer phase. Apart from a significant chess exercise effect, F (3, 204) = 13.26, MSE = 244.70, $p < .001$, a significant group effect of JOL, F (1, 68) = 6.38, MSE = 535.35, $p < .05$, was also observed. Post-hoc tests showed that the groups that did not provide JOLs had higher transfer scores than the groups that did provide JOLs, $p < .05$. The group effect of move selection was nearly significant, F (1, 68) = 3.60, MSE = 535.35, $p = .06$ (see Table 1).

In a similar repeated measures analysis on the restudy predictions, a significant effect of JOL, F (1, 27) = 5.94, MSE = 613.93, $p < .05$, move selection, F (1, 27) = 18.83, MSE = 613.93, $p < .001$, and a significant interaction between JOL and move selection, F (1, 27) = 6.33, MSE = 613.93, $p < .05$, were observed. The No JOL groups performed better than the JOL groups, and the forced move selection groups performed better than the free move selection groups. The interaction effect indicated that the forced groups did not differ in restudy predictions between the JOL and No JOL instruction, whereas the free selection groups performed better in the No JOL instruction than in the JOL instruction. The effect of chess exercise was not significant, $F < 1$.

Table 1
Mean percentage of correct predictions in the prediction phase, the transfer phase, and the restudy phase per group. Standard deviations between brackets

Condition		Prediction phase	Transfer phase	Restudy phase
JOL instruction	Free move selection	43.3 (12.3)	47.5 (11.1)	32.5 (32.0)
	Forced move selection	52.1 (7.3)	54.7 (12.7)	56.8 (14.0)
	Total	47.7 (10.9)	51.1 (12.3)	46.6 (25.9)
No JOL instruction	Free move selection	45.1 (6.7)	55.9 (9.8)	45.6 (19.4)
	Forced move selection	56.3 (12.8)	59.8 (12.3)	59.1 (15.1)
	Total	50.5 (11.5)	57.8 (11.1)	52.0 (18.6)
Total	Free move selection	44.2 (9.7)	51.8 (11.1)	40.3 (25.6)
	Forced move selection	54.2 (10.5)	57.3 (12.6)	57.9 (14.4)
	Total	49.1 (11.2)	54.5 (12.1)	49.5 (22.2)

Correlations between JOLs, transfer predictions, and move selections
To assess participants' monitoring accuracy, we computed the correlation between judgments of learning and move predictions for the groups that provided JOLs (i.e., the JOL free and JOL forced group). When this correlation was positive and significant, this indicated adequate monitoring accuracy, because participants' JOLs and their performance agreed. Furthermore, we examined participants' degree of self-regulation by correlating JOLs with move selections for restudy. A significant, negative correlation indicates accurate self-regulation, because participants chose those moves for restudy that they considered most difficult. Because of the low number of studied chess moves per individual, we computed correlations per group instead of per individual, in which rows represented separate moves. Since the transfer phase preceded the move selection phase and followed the JOL phase, it is possible that participants based their move selections for restudy partially on information from the transfer phase. To analyze this possibility, we also calculated the correlation between transfer predictions and move selections.

The correlations between JOLs and move selections, between JOLs and move predictions, and between transfer predictions and move selections were not significant for the JOL free selection group. However, for the JOL forced group, the mean correlation between JOLs and predictions was .25 ($p < .05$), indicating that this group was to a certain extent able to assess their understanding of the moves. The correlation between JOLs and move selections was -.31 ($p < .01$), which shows that participants in this group based their move selections in part on which moves they thought they did not understand. The correlation between transfer predictions and move selections was not significant (r = -.15, p > .05). More detailed analysis of the development of monitoring accuracy and self-regulatory behavior showed that for the JOL forced group correlations were low at the start of the learning phase, but increased over the course of the experiment (from -.03 to -.26 for monitoring accuracy, and from .18 to .29 for self-regulation). This indicates that, possibly because of the absence of prior knowledge, participants needed several practice cycles to learn to estimate their understanding of the chess exercises and to learn to base their move selections on this estimation. In general, monitoring accuracy and self-regulation correlations were fairly low.

Table 2

Mean correlations over the four chess exercises between JOLs, predictions, transfer predictions, and move selections per group. These correlations are based on the total added values per move, ignoring the participant level

	Correlation		
Group	*JOL & predictions*	*JOL & move selections*	*Transfer predictions & move selections*
JOL free selection	.07	-.09	-.11
No JOL free selection	-	-	.01
JOL forced selection	.25*	-.31**	-.15
No JOL forced selection	-	-	-.14

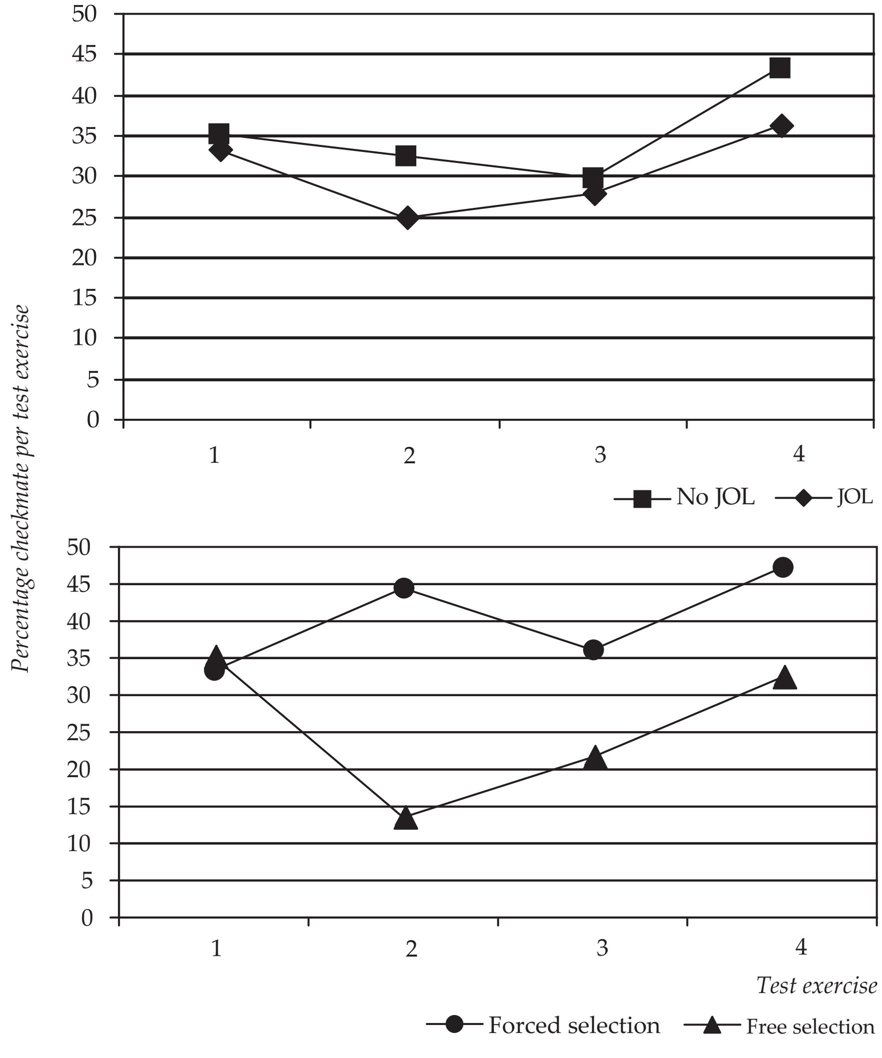
Note. * p < .05** p < .01

Test phase

In the test phase, participants in all conditions had to play four endgames against the computer and checkmate the black King. Repeated measures ANOVA with test exercise (1, 2, 3, 4) as a within-subject factor, and JOL (present or absent) and move selection (free or forced) as the between-subject factors revealed that the test exercise effect was not significant, F < 1. The effect of move selection (free or forced) was significant, F (1, 68) = 6.89, MSE = 0.36, p < .05, but the effect of JOL (present or absent) was not, F < 1. Post-hoc analysis revealed that the forced move selection groups performed better on the test exercises than the free move selection groups. The estimated marginal mean percentage of checkmates for the JOL free selection condition was 23.6% (SD = 27.7), for the No JOL free selection condition it was 27.6% (SD = 27.5), for the JOL forced selection condition it was 37.5% (SD = 32.3), and for the No JOL forced selection it was 43.1% (SD = 33.0). In Figure 5, these data are represented as a function of the between-subjects factor JOL (present or absent) and move selection (free or forced).

Figure 5

Mean percentage checkmate per test exercise per group



Discussion

This study examined the effect of providing judgments of learning and selecting material for restudy on learning to play chess. Participants had to learn an endgame of chess by studying several chess exercises provided by a chess computer. While studying these exercises, participants were instructed to predict and explain the moves of the computer. In a four groups 2 x 2 design, we varied both the JOL instruction by asking two of the four groups to provide JOLs, whereas the two other groups did not provide JOLs. Moreover, we varied the move selection instruction, by asking two of the four groups to select at least two moves for restudy and leave the two other groups free in the number of move selections. In the test phase, participants played four versions of the endgame against the computer.

The results indicate that in the learning phase, participants who were forced to select at least two moves for restudy performed better when predicting the computer moves, even when controlling for the fact that they had more practice because of the higher number of move restudies. Moreover, the groups that had to provide JOLs performed worse than the groups that did not provide JOLs. In the transfer prediction task, a similar pattern was observed: The groups that provided JOLs performed worse than the groups that did not provide JOLs. The difference in performance between the forced and free move selection groups was almost significant, in favor of the groups that had to select at least two moves for restudy. In the test phase, no differences were observed between the groups that did or did not provide JOLs. However, the groups that were forced to select at least two moves outperformed the groups that were free in number of move selections. These findings show that to foster learning to play chess, the move selection and restudy instruction was more effective than the instruction to estimate degree of understanding. It appears that asking novices to judge their understanding when learning to play chess has only a small effect on self-regulation and performance, and might even have a detrimental effect on learning given the higher performance of the No JOL groups compared to the JOL groups in all prediction tasks in the learning phase.

The correlational analysis revealed that the JOL forced selection group had better metacognitive monitoring and self-regulation compared to the JOL free selection group: Not only were they aware that more selections were necessary to learn the endgame, but the JOLs they provided also correlated positively with the predictions they had made. Above all, their move selections correlated negatively with the level of their JOLs, which indicates that, in contrast to the JOL free selection group, their selections were based on their metacognitive estimation of which moves needed restudy. However, the effect of providing JOLs was not as large as expected. For instance, the monitoring accuracy and self-regulation correlations, although significant, were low. Moreover, the JOL groups performed lower on all prediction tasks in the learning phase than the No JOL groups and did not differ from the No JOL groups in the test phase. Taking these findings into account, it is likely that the general advantage of forcing novice learners to provide JOLs seems limited. Considering the absence of prior knowledge, it is possible that the JOL instruction alone was not sufficient to support learners to adequately estimate their degree of understanding of the endgame. Apparently, for novice chess players, monitoring the accuracy of their predictions and self-regulating their study efforts based on that information is highly complex and requires practice to develop.

These results can be interpreted within the framework of the expertise reversal effect (e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003). A learner's prior knowledge level determines largely what information is relevant and irrelevant during learning. If a learner has acquired schemas on a specific topic, these can allow the learner to avoid extensive processing of information and thereby prevent an overload of working memory (Kalyuga et al., 2003). However, inexperienced learners are forced to process the incoming information without the help of pre-organized schemas that aid in classifying the information. The instruction they receive should be directed at minimizing working memory load of irrelevant information in order to optimize learning of the relevant procedures (Sweller, 1999; Sweller, Van Merriënboer, & Paas, 1998). For example, a study by Tuovinen and Sweller (1999) showed that learners without previous experience in database management learned more from instruction on how to use a database that incorporated worked examples, than from exploratory practice. Worked examples were found to effectively reduce ineffective cognitive load and enhance learning (e.g., Carroll, 1994; Paas & Van Merriënboer, 1994), whereas exploratory learning, to a certain extent comparable to the instruction in the present study, is seen as placing a high demand on working memory. For learners with previous database experience the type of practice made no difference to their learning, because of pre-existing schemas that helped them organize the information. In the present study, the JOL procedure required learners to estimate how confident they were that they would correctly predict the computer moves. Since participants had no prior chess knowledge, and therefore no prerequisite relevant schemas, this instruction possibly placed a high load on working memory, leading to decreased learning. The fact that the No JOL groups performed better than the JOL groups, especially in the transfer prediction task underlines this possibility. The JOL procedure preceded the transfer phase, which might have caused reduced processing attention in the JOL groups, leading to lower performance in the transfer phase. Why the forced selection groups performed better than the free selection groups, even when controlling for number of move selections actually made, is unclear. One possibility is that the forced selection instruction aided in directing attention to efficient move selection for restudy, which might have led to a lower mental load than the free selection instruction. Therefore, mental load was freed to efficiently select those moves for restudy that needed restudying. To reduce working memory load associated with the JOL instruction, it would be interesting to integrate the JOL instruction with the prediction instruction in future research, instead of separating it as in the present study.

The present findings suggest that the JOL paradigm can not be applied to novice learning without further ado. It is possible that metacognitive activities as self-monitoring and self-regulating require valuable processing capacity, and thus do not prove effective until more advanced levels of expertise have been attained. Research is needed that incorporates groups of different levels of expertise to further unravel the possible effects of providing JOLs and selecting material for restudy on cognitive skill acquisition.

Chapter 6

Improving metacomprehension accuracy and self-regulation when learning to play a chess endgame: The effect of a generation task and performance feedback[5]

[5] This chapter is submitted for publication as: De Bruin, A.B.H., Rikers, R.M.J.P., & Schmidt, H.G. (2006). *Improving metacomprehension accuracy and self-regulation when learning to play a chess endgame: The effect of a generation task and performance feedback*. Manuscript submitted for publication.

Abstract

The positive learning effect of metacognitive strategy instructions based on the cue-utilization framework (Koriat, 1997) has been shown in memorizing word pairs and studying expository text. The present study explored the effect of these types of instructions in cognitive skill acquisition on learners with low prior knowledge. In Experiment 1, the quality of metacomprehension and self-regulation of novices learning to play chess was compared to that of more experienced chess players. The low spontaneous use of these metacognitive activities, even in more experienced chess players, indicated the need for more guided instruction. Therefore, Experiments 2 and 3 attempted to enhance these activities by stimulating learners to generate cues that were predictive of future performance prior to monitoring comprehension. The results indicated that the generation instruction positively affected metacomprehension accuracy and transfer performance, but not self-regulation. Moreover, providing learners with performance feedback had no influence on metacognitive activities, but did improve transfer performance. This study showed that a generation instruction directed at increasing metacomprehension accuracy in low prior knowledge learners on a complex task had a positive effect on transfer performance. The appropriateness of restudy activities in cognitive skill acquisition, and the need for transfer tasks that assess understanding at the long term are further discussed.

Over the past two decades, research on metacognition has been directed at uncovering the key strategies that self-regulated learners apply when approaching study tasks. For example, self-regulated learners set clear and attainable goals, monitor learning activities, and adjust these when the monitoring process identifies deviations from attainment of these goals (Winne, 1995a; 2001; Zimmerman & Schunk, 2001). Given the evident positive effect of self-regulated learning on achievement (e.g., Zimmerman & Schunk, 2001), research is required that uncovers which factors determine metacognitive activities, and how these can be improved in learners with low metacognitive monitoring. Theoretical models that explain self-regulated study can be described as discrepancy-reduction models (e.g., Butler & Winne, 1995; Dunlosky & Hertzog, 1997; Thiede & Dunlosky, 1999). That is, the individual has a desired level of understanding of the to-be-studied material, and monitors during learning to what extent this state has been reached. If a discrepancy is identified between desired and current level of understanding, the learner will decide to continue studying the material. During restudy, the learner will again monitor the level of understanding and determine whether further restudying is necessary. This process continues until the discrepancy between desired and current level of understanding has disappeared.

Evidence for this explanation is mainly found in research on memorizing word pairs. For example, Son and Metcalfe (2000) list a number of studies that demonstrate that if a learner is able to determine which material is understood, and which is not, more study time will be allocated to the least-understood material. Moreover, Thiede (1999) showed that the level of monitoring accuracy determines the effectiveness of self-regulation of study: Learners who more accurately monitored learning selected material that needed restudying more appropriately, and performed better on a subsequent criterion test.

In the word-pair paradigm, participants are typically asked to study a list of translation pairs (i.e., Swahili-English translations; Nelson, Dunlosky, Graf, & Narens, 1990). Afterwards, they are presented with the first word of each pair and asked to estimate the chance of retrieving the second word on a subsequent test. These estimations (ranging from 0% to 100%) are termed "judgments of learning" (JOLs). The higher the JOL, the more confident the learner is to retrieve the correct word on the test. Accurate monitoring of learning is indicated by a highly positive correlation between JOLs and test performance: The learner was able to assign high JOLs to words that were recalled, and low JOLs to words that were not recalled. After the JOL procedure, learners are asked to select the word pairs they wish to restudy. Adequate self-regulation is characterized by a highly negative correlation between JOLs and selection of word pairs for restudy: word pairs that were assigned high JOLs were not selected for restudy, whereas word pairs with low JOLs were selected. Note that adequate self-regulation of study is dependent on accurate metacognitive monitoring: If learners assign incorrect JOLs to word pairs (e.g., high JOLs to word pairs that were not correctly retrieved), they will not select the word pairs that in fact need restudying.

The positive relation between JOLs and subsequent performance, and the negative relation between JOLs and item selection has been established repeatedly using the word-pair paradigm (e.g., Cull & Zechmeister, 1994; Dunlosky & Hertzog, 1997; Nelson et al., 1994; Thiede & Dunlosky, 1999). Koriat (1997) argues that, when providing JOLs, learners use certain cues that are indicative of ease or difficulty of learning the word pairs. Within this cue-utilization view, learners are thought to apply specific rules or heuristics to arrive at an assessment of the probability of recalling the item later. Although the evidence for the cue-utilization approach has provided insight into the cognitive mechanisms that explain why JOLs predict memory performance, research in this field has generally led to few educational implications. Given that the relevance of memorizing word pairs only applies to a small subset of the educational contexts, research is needed that addresses the validity of the cue-utilization framework in educationally more relevant situations. In that regard, research on studying expository text has shown that, when preceded by a keyword or summarization instruction, learners are able to accurately assess metacomprehension[†] of the texts, and regulate further study behavior by using that information (Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). However, despite the possibly high educational relevance, the applicability of the cue-utilization view has not been tested in the field of cognitive skill acquisition. In contrast to research on word pairs, studies in cognitive skill acquisition that involve straightforward metacognitive strategies potentially result in educational implications that are transferable to diverse skill domains. Therefore, the field is in need of research that examines the applicability of the cue-utilization view in cognitive skills.

Besides the need for research on metacognition in educationally more relevant contexts, a second theme that requires attention concerns the influence of learner expertise on metacognitive strategy training. More specifically, the question is to what extent learners of low prior knowledge are able to benefit from training in specific metacognitive skills. For example, in cognitive skill acquisition, certain characteristics can impede self-regulated learning in novices. Due to the absence of previous experience,

[†] The term used to denote the predictive accuracy of comprehension ratings in learning from text is “metacomprehension accuracy”, whereas the term used in learning word pairs is “monitoring accuracy” (Maki, 1998). In chess, we also prefer the term “metacomprehension accuracy”, since this more clearly emphasizes that learners reflect on how well study material is understood.

none of the subskills have been automated, which causes execution of the skill to pose a high demand on working memory (Winne, 1995a). Requiring learners, moreover, to self-monitor and self-regulate their learning could induce a working memory overload. Therefore, Winne (1995a) argues that self-regulated learning should be delayed until individuals have at least proceduralized the rules of the domain. On the other hand, although metacognitive strategies are traditionally associated with expert learners, there is evidence that metacognitive strategies can be taught effectively to inexperienced learners. Azevedo and Cromley (2004) provided participants with a 30-minute training on how to use self-regulated learning before studying a hypermedia environment that explained the circulatory system. Even though the learners possessed low prior knowledge of the domain, those who received the self-regulated learning training experienced a larger shift in their mental models than those who did not. Furthermore, a study by Butler (1998) revealed that self-regulation skills of poor readers could be successfully improved through extensive training. Students were even able to transfer the learned skills to different contexts. These studies underscore the potential of low prior knowledge learners to benefit from metacognitive strategy instruction. However, given the extensiveness of the metacognitive training they used, identifying facets of the training that caused the learning effect is difficult, and transfer to other educational domains at this point not possible. Research is needed that more specifically manipulates the content of the metacognitive strategy instruction, and studies its effect on cognitive skill acquisition in novices, thereby allowing for the possibility of formulating general educational implications.

The present study was designed to provide insight into how novices' metacognitive monitoring and self-regulation during cognitive skill acquisition can be improved. More specifically, two research questions were addressed. First, we examined to what extent novice learners can benefit from metacognitive strategy training compared to more experienced learners. Therefore, in Experiment 1, we compared the metacognitive activities of novice and more experienced chess players, to assess to what extent accurate metacognitive monitoring and self-regulation is dependent on previous task experience.

Our second research question concerned examining how metacognitive monitoring and self-regulation can be improved in novice learners, and to what extent improved metacognition affects performance and transfer of learned skills. Therefore, experiments 2 and 3 tested the effect of specific monitoring and self-regulation instructions in novices. These instructions were designed from the perspective of the cue-utilization framework (Koriat, 1997). Until now, evidence for the cue-utilization framework mainly comes from studying word pairs or text in knowledgeable participants. The current experiments allowed us to evaluate novices' ability to profit from monitoring and self-regulation instruction, and to examine how novices' metacognitive activities can be stimulated in chess skill acquisition.

Experiment 1

The goal of the first experiment was to examine to what extent low prior knowledge learners are able to benefit from self-regulatory strategy training in chess skill acquisition. According to research by Kanfer and Ackerman (1989) and Winne (1995a), under the high working memory demands of non-automatized skills, the execution of self-regulated learning strategies will not exert their effect until learners have proceduralized the rules of a domain. However, some studies (e.g., Azevedo & Cromley, 2004; Butler, 1998) have shown that self-regulatory training can positively affect learning in low prior knowledge learners. Given that these studies made use of a multifaceted, extensive self-regulation training, they do not provide detailed information on which metacognitive activities specifically foster learning in novices. In this study, we compared novices and more experienced chess players on two specific forms of metacognitive activities, namely metacomprehension accuracy and self-regulation when learning to play the endgame of King and Queen against King. We hypothesized that the more experienced chess players would not only learn to play the endgame better, but would also more accurately monitoring understanding and regulate their study behavior.

Method

Participants. Participants were 38 undergraduate students from Erasmus University Rotterdam, The Netherlands (*M* age = 20.5, *SD* = 1.8). Eighteen participants had never played chess before, nor were aware of the rules of chess (i.e., "inexperienced chess players"‡). The remaining twenty participants were selected based on their level of chess experience (i.e., "experienced chess players"). These were participants who had proceduralized the chess rules and had in previous years played chess, but infrequently. The selection of the experienced chess players contained three criteria. First, participants had to obtain a perfect score on a 9-item true-false test on the basic rules of chess. What is meant by the basic rules of chess is explained below. Second, participants had to play three examples of the endgame of King and Queen

‡ This group is the same group as the "JOL forced selection condition" in De Bruin, Rikers, & Schmidt, 2005a.

against King (the KQK endgame) against a chess computer. Of these three endgames, participants had to be able to solve the easiest one, which was an example of a game of checkmate in one move. Of the remaining two (checkmate in three and nine moves) participants had to solve exactly one, but never in the minimum number of moves, to guarantee that they did not master the endgame. Finally, participants were asked about their chess activities in the past. Only participants who had never joined a chess club were allowed to take part in the experiment. Of the 107 participants tested, 20 met all three criteria and took part in the study.

Materials. At the start of the experiment, the inexperienced chess players studied a computer presentation about the basic chess rules (i.e., what are the legal moves for King and Queen? What is capturing, check, checkmate, and stalemate?). These rules can be clearly distinguished from the principles that underlie the endgame: Knowing what the legal moves of a Queen and King are does not provide enough information for a learner to play this endgame correctly. Afterwards, participants completed the same true-false test that the experienced players had taken in the selection procedure, to assess whether they had understood the rules.

The chess computer program "Fritz 7" provided the chess exercises for the learning phase and the test phase, which consisted of examples of the KQK endgame. This chess program plays at Grandmaster level (World Chess Federation rating of approximately 2700). Since learning to play chess normally takes several study sessions, a subpart of the game was selected that could be studied in isolation (the KQK endgame). The difficulty level of the chess exercises in the learning phase was constant: All exercises were examples of endgames of checkmate in six moves, representing the same underlying chess principles. These chess exercises were presented on a computer screen, embedded in a computer application that recorded participants' responses. In the test phase, participants played four versions of the endgame against the computer.

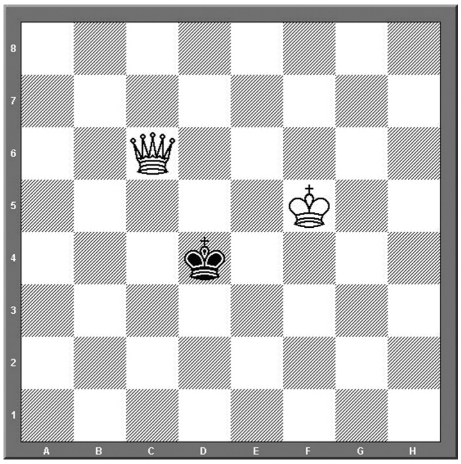
Procedure. The experiment consisted of three phases: the basic chess rules phase, the learning phase, and the test phase. Before the start of the learning phase, the inexperienced participants studied the presentation about the basic chess rules in a self-paced way. This took on average about 25 minutes. Afterwards, they completed the paper and pencil test. Any errors were explained by the experimenter before proceeding to the learning phase. At that point, the inexperienced players had not proceduralized the rules, but were only able to apply them in a non-automatic way.

The goal of the learning phase was to have participants monitor their comprehension and self-regulate their study behavior of the KQK endgame.

This phase had an identical structure for both experienced and inexperienced chess players. However, before being able to monitor understanding of the KQK endgame, participants needed to become familiar with the endgame. Therefore, participants first studied an example of the KQK endgame by predicting the moves for white, and evaluating their prediction by observing the computer's moves afterwards. That is, participants predicted move 1 for white, then observed the actual computer move 1 for white, then saw the computer's move 1 for black, and then predicted move 2 for white, etc. until the game ended after six moves. Participants were given 40 seconds per prediction. Because in this endgame black only consists of a King, white (i.e., King and Queen) is, in theory, able to win. Therefore, white's task is to correctly play and win the endgame according to the endgame's principles. These principles were nowhere stated, but had to be inferred by studying and predicting the computer moves. This part of the learning phase was termed the prediction phase. An example of a chess position used in the learning phase is provided in Figure 1.

Figure 1

Example of a chess position used in the learning phase



To promote thorough processing, we instructed participants in the prediction phase of the experiment to provide self-explanations while predicting the next move (Chi, De Leeuw, Chiu, & LaVancher, 1994; De Bruin, Rikers, & Schmidt, in press; Renkl, 1997). Since the effect of self-explaining on learning to play chess is described elsewhere (De Bruin et al., in press), the self-explanations were not further analyzed in this study.

After all moves of an endgame had been predicted in the prediction phase, participants proceeded to the Judgment of Learning phase (JOL phase). This phase allowed learners to monitor their understanding of the previously studied endgame. Participants were shown all positions from the previously studied endgame at which white was about to make a move. These positions were presented in random order, to prevent that learners could restudy the sequence of moves when formulating JOLs. For every move, participants were asked to judge how well they had understood the move that had to be made by pressing a corresponding button: “How confident are you that you will predict a similar move correctly in the future?” (0% = definitely will not be able to predict, 20%, 40%, 60%, 80%, and 100% = definitely will be able to predict). No self-explanations were generated during this phase.

When participants had provided JOLs for all six moves, they were shown the same chess exercise, but now in mirror image position, and were once more asked to predict the moves white would make. This reprediction phase allowed participants to assess the accuracy of the JOLs, and enabled us to calculate a monitoring accuracy correlation between JOLs and repredictions. Using a mirror image position, the content of the exercise remained identical, but the appearance changed. Therefore, we ensured that participants had to apply the correct principle, instead of basing it on memorization. Since this phase of the experiment was not intended to infer endgame principles, participants did not self-explain and were given 20 instead of 40 seconds to predict each move.

Finally, in the move selection phase, which was designed to trigger self-regulation in the learners, participants were allowed to select and restudy the moves of the mirror image exercise. Participants were shown the positions of the mirror image exercise at which white was about to move, and were asked to press “y” if they wanted to restudy another example of this move, and “n” if not. No self-explanations were provided during this phase. Participants were instructed to select at least two moves for restudy. The restudy phase consisted of studying the selected moves, but now of another mirror image of the same exercise, under the same instructions as in the prediction phase. After restudying all selected moves, a new chess exercise of the same difficulty level was presented. They were again asked to first predict the moves of the computer, then provide JOLs, assess JOL accuracy in the reprediction phase, and finally select and restudy at least two moves. Participants studied four chess exercises under these instructions. They worked through the application individually, which lasted on average an hour. An overview of the procedure of the learning phase for all three experiments, comparable to Thiede et al.’s (2005) description is found in Table 1.

Table 1

Overview of the experimental procedures of the learning phase in Experiments 1 to 3. The learning phase consisted of four (Exp 1 and 2) or five (Exp 3) chess exercises studied under these procedures. The generation task group in Experiment 2 and the No-feedback group in Experiment 3 are similar, except for the fifth chess exercise used in the no-feedback group in Experiment 3.

	Experiment 1 Experienced versus inexperienced chess players				
Inexperienced (N = 18) and experienced chess players (N = 20)	P1, P2 .. P6	J1, J2 ..J6	R1, R2 .. R6	S1, S2 .. 6	RS1, RS2 … RS6
	Experiment 2 Generation task versus no generation task				
Generation task group (N = 25)	P1, P2 .. P6	G1, J1, S1, G2, J2, S2, … G6, J6, S6			RS1, RS2 … RS6
No-generation task group (N = 25)	P1, P2 .. P6	C1, J1, S1, C2, J2, S2, … C6, J6, S6			RS1, RS2 … RS6
	Experiment 3 Feedback versus no feedback				
Feedback group (N = 23)	P1, P2 .. P6	R1, J1, F1, S1, R2, J2, F2, S2, … R6, J6, F6, S6			RS1, RS2 … RS6
No-feedback group (N = 20)	P1, P2 .. P6	R1, J1, S1, R2, J2, S2, … R6, J6, S6			RS1, RS2 … RS6

Note. P1 = participants predicted move 1; J1 = Participants judged comprehension of move 1; R1 = Participants repredicted move 1; S1 = Participants selected move 1 for restudy; RS1 = Participants restudied move 1; G1 = Participants predicted move 1 to generate cues prior to JOL; C1 = Participants counted the number of moves white could make at move 1.

After the learning phase had ended, participants were instructed to play four new KQK endgames against the computer and checkmate the black King in as few moves as possible. The first two of these were comparable in difficulty to the learning phase (checkmate in six moves), whereas the final two were slightly more difficult (checkmate in eight and nine moves). Since all test exercises required application of the same principles, they were analyzed as a whole. Following the rules of the World Chess Federation (FIDE), a game ended in a draw when either 50 moves had been made and no piece was captured, the black King had taken the white Queen (checkmate impossible), stalemate had occurred, or the same position had appeared three times in the same game. Participants received maximally five minutes per test exercise.

Analysis. Participants' moves in the prediction, the reprediction, and the restudy phase of each chess exercise were compared to the computer's optimal move. When a predicted move was equal to or as good as the computer move (sometimes several equally good moves were possible) two points were awarded. Predicted moves that led to a game situation that needed one extra move to checkmate the black King than the optimal move were scored as one. All other predictions (i.e., leading to situations that required two or more extra moves to checkmate the King) were scored zero. Maximally twelve points per chess exercise were awarded. The prediction scores are represented as percentages of the maximum score of twelve points.

Repeated measures analyses of variance were used to analyze the data in the learning phase. First, in a 4 x 2 analysis, the difference between the inexperienced and experienced chess players in number of moves selected for restudy was examined, with chess exercise (first, second, third, fourth) as the within-subjects factor, and expertise level (inexperienced or experienced) as the between-subjects factor. Similar repeated measures ANOVAs were used to analyze the difference in accuracy of move predictions in the prediction, reprediction, and restudy phase. The results of the test exercises were examined using a oneway analysis of variance with expertise level as the between-groups factor and the number of times the participants checkmated the black King (ranging from 0 to 4) as the dependent variable. Interaction effects were calculated, but will only be mentioned when significant. All alpha levels were set at .05.

The two main indicators of metacognitive activity in this and the following experiments were metacomprehension accuracy and self-regulation. Since a number of participants had no variance in JOLs and move selections (i.e., they always selected two moves), calculating Goodman-Kruskal gamma correlations (as is common in research on metacognition: e.g., Maki & Serra, 1992; Nelson, 1984; Thiede et al., 2003) would lead to discarding data of several participants. This would threaten the generalizability of the results for the complete sample. Instead, to determine metacomprehension accuracy, we calculated separately for the group of inexperienced and experienced chess players and per chess exercise what the mean JOL score was, and what the mean reprediction score was, and combined these, to arrive at one correlation per exercise and per expertise level. To evaluate participants' level of self-regulation, bivariate correlations were calculated between mean JOLs and mean number of move selections per chess exercise and per expertise level. Although we did not compare JOLs with specific move selections, this procedure allowed us to examine the mean metacomprehension accuracy and self-regulation of the participant groups per chess exercise.

Results

Learning phase. First, we analyzed the number of move selections made by the experienced and inexperienced players in the learning phase. A repeated measures analysis of variance with chess exercise (first, second, third, and fourth) as the within-subjects factor, and expertise level (inexperienced or experienced) as the between-subjects factor, showed a main effect of expertise level, F (1, 36) = 4.84, MSE = 5.47, $p < .05$, $\eta^2_p = .12$, which indicated that inexperienced chess players selected on average more chess moves for restudy. Moreover, an effect of chess exercise was found, F (1, 36) = 16.38, MSE = 1.95, $p < .001$, $\eta^2_p = .31$. The number of move selections decreased over the four exercises.

To examine the accuracy of the move predictions in the chess exercises, repeated measures analyses of variance were performed on the prediction scores in the prediction, reprediction, and restudy phase with chess exercise as the within-subjects factor, and expertise level as the between-subjects factor. For the prediction phase, a significant effect of group was found, F (1, 36) = 14.39, MSE = 345.78, $p < .01$, $\eta^2_p = .29$, which indicated that the experienced chess players had higher prediction scores than the inexperienced chess players. Also, an effect of chess exercise was observed, F (1, 36) = 134.00, MSE = 241.53, $p < .001$, $\eta^2_p = .79$, showing that prediction scores improved over the course of the chess exercises. The same pattern was observed for the reprediction scores: A significant effect of expertise level in favor of the experienced chess players, F (1, 36) = 39.47, MSE = 497.10, $p < .001$, $\eta^2_p = .52$, and a significant effect of chess exercise was found, F (1, 36) = 27.86, MSE = 353.82, $p < .001$, $\eta^2_p = .44$. The restudy scores showed the same effects: An effect of expertise, F (1, 28) = 8.44, MSE = 529.30, $p < .01$, $\eta^2_p = .23$, and an effect of chess exercise in favor of experienced chess players, F (1, 28) = 36.60, MSE = 496.86, $p < .001$, $\eta^2_p = .57$. However, since the latter analysis was based on the subset of participants who selected moves on all four exercises, results should be interpreted with caution. An overview of the prediction scores in the learning phase for all three experiments is provided in Table 2.

Table 2

Results of the learning phase on various dependent variables for the three experiments. For Experiment 1, JOLs were expressed as percentages, for Experiments 2 and 3 on a scale from 1 to 7. Standard deviations between brackets

Group	Mean move selections	Mean JOLs	Prediction score	Reprediction score	Restudy score
			Experiment 1		
Inexperienced chess players	3.36 (1.42)	41.48 (12.22)	51.12 (7.20)	54.33 (13.08)	56.77 (14.01)
Experienced chess players	2.59 (0.86)	63.42 (14.15)	63.60 (11.08)	77.63 (9.79)	75.60 (11.83)
Total	2.95 (1.21)	53.02 (17.16)	58.11 (10.85)	66.72 (15.92)	66.68 (15.90)
			Experiment 2		
Generation task group	2.06 (1.07)	4.52 (0.98)	51.13 (15.41)	58.64 (17.31)	75.65 (12.02)
No-generation task group	1.57 (0.74)	4.60 (0.92)	51.30 (11.97)	-	71.66 (18.92)
Total	1.82 (0.94)	4.53 (0.94)	51.22 (13.65)	58.64 (17.31)	73.65 (15.81)
			Experiment 3		
Feedback group	1.72 (0.82)	4.84 (1.06)	58.91 (14.35)	63.26 (17.01)	46.04 (19.28)
No-feedback group	2.11 (1.58)	4.43 (1.03)	52.50 (9.47)	54.08 (14.29)	38.80 (21.07)
Total	1.88 (1.22)	4.68 (1.04)	56.26 (12.46)	59.59 (16.51)	43.03 (20.25)

Metacomprehension accuracy was calculated by first determining the mean JOL and the mean reprediction score for each chess exercise and per individual. Then, the correlation between these two variables was determined separately for the experienced and inexperienced chess players. This same procedure was applied to JOL and number of move selections to provide a measure of self-regulation. Since the correlations were computed per group, instead of per individual, statistical analysis of the differences in correlations between groups was not possible. Therefore, interpretation of the differences in these correlations between groups should be performed with caution. These correlations are shown in Table 3. As shown, the metacomprehension accuracy correlations are low for the inexperienced chess players (mean r = .14) compared to the experienced group (mean r = .38). The self-regulation correlations are on average higher for the inexperienced chess players (mean r = -.40) than for the experienced chess players (mean r = -.20). That is, the moves that the inexperienced chess players selected for restudy were moves that they had awarded low JOLs. The discrepancy between lower metacomprehension accuracy and higher self-regulation of the inexperienced chess players can be explained by the fact that their move selections were not always the correct choices. They represented the moves learners *believed* they had difficulty with, rather than the moves they *in fact* had difficulty with.

Table 3

Monitoring accuracy and self-regulation correlations for experienced and inexperienced chess players in Experiment 1. Monitoring accuracy is expressed through the correlation between JOLs and reprediction scores. Self-regulation is expressed through the correlation between JOLs and move selections. P values between parentheses

Expertise level	Monitoring accuracy			
	Exercise 1	*Exercise 2*	*Exercise3*	*Exercise 4*
Inexperienced chess players	.09 (.73)	.10 (.68)	.01 (.96)	.35 (.16)
Experienced chess players	.33 (.15)	.54 (.01)	.29 (.22)	.36 (.13)
	Self-regulation			
	Exercise 1	*Exercise 2*	*Exercise 3*	*Exercise 4*
Inexperienced chess players	-.50 (.04)	-.47 (.05)	-.45 (.06)	-.21 (.41)
Experienced chess players	-.10 (.68)	-.33 (.15)	-.21 (.38)	-.15 (.53)

Test phase. For the test phase, we calculated the number of times the learners checkmated the black King. Scores ranged from 0 (no checkmates) to 4 (all test exercises checkmate). A oneway analysis of variance revealed that the experienced chess players checkmated the black King more often than the inexperienced chess players, F (1, 36) = 17.03, MSE = 1.26, $p < .001$, $\eta^2_p = .32$. On average, the experienced chess players checkmated the black King 2.95 times (SD = 1.00), and the inexperienced chess players 1.44 times (SD = 1.25).

Discussion

As expected, the experienced chess players outperformed the inexperienced chess players in learning scores and test performance. However, the lack of interaction between performance and expertise level in the learning phase indicates that both groups developed understanding of the endgame at the same rate. As to metacomprehension accuracy, the correlations between predictions and JOLs were on average higher for the experienced chess players. Furthermore, the self-regulation correlations were higher for the inexperienced chess players. However, given the inaccuracy of the JOLs of this group, their selections of moves for restudy, although related to JOLs, were not accurate. These move selections corresponded only to a small extent to the moves the learners in fact had predicted poorly.

In sum, the accuracy of metacognitive activities does not directly improve as a result of having proceduralized the rules of chess, as Winne (1995a) assumes. Other factors need to be identified that explain the low metacomprehension accuracy and self-regulation when learning to play chess. According to prior research (Koriat, 1997; Thiede et al., 2003, 2005), metacomprehension accuracy depends on the activation of cues that can be used to judge comprehension. For these cues to have an effect, it is important that they are active at the time of judgment. In his cue-utilization framework, Koriat (1997) describes how mnemonic cues that are generated as the result of encoding operations of the learner improve metamemory. These cues indicate, for instance, the ease with which the word pairs were processed (Shaw & Craik, 1989), or experienced familiarity of the items (Kelley & Lindsay, 1993). The availability of these cues at time of judgment can be experimentally manipulated. For example, asking learners to generate keywords of a studied text, prior to rating comprehension, led to more accurate comprehension ratings and more effective self-regulation. Learners who generated keywords more often selected the correct text for restudy than learners who did not generate keywords (Thiede et al., 2003). Thiede and colleagues explain the keyword effect in terms of access to the situation model of the text (Kintsch, 1988). The situation model provides an accurate predictor of long-term memory of the text, and, therefore, using information from the situation model when monitoring comprehension improves accuracy. In more general terms, the keyword effect can be explained within the cue-utilization framework (Koriat, 1997) as the activation of cues that are predictive of future performance prior to rating comprehension.

In Experiment 1, learners were not explicitly stimulated to activate cues that are predictive of future performance when providing JOLs, but had to activate these cues for themselves. Especially for the inexperienced chess players, a failure to activate cues might provide an explanation for the low metacomprehension accuracy. Therefore, we decided to insert a cue-generation task before learners were asked to provide JOLs in Experiment 2. That is, after having predicted the moves of a chess exercise, participants were shown a mirror image of the same chess position and were once more asked to predict the next computer move. This prediction was hypothesized to lead to the generation of cues that are indicative of understanding. After every prediction, participants were immediately asked to provide a JOL for that move. The immediacy ensured that the generated cues were still available. The control group did not predict the next move, but only counted the number of moves white could legally make. Since the generation task activated cues that were predictive of future performance and that were available at the time of JOL, we hypothesized that this group would show better metacomprehension accuracy and self-regulation, and that they would outperform the control group in the test phase. Finally, the self-explanation instruction in the learning phase might have interfered with the metacognitive activities participants engaged in. Therefore, in Experiment 2, participants did not provide self-explanations.

Experiment 2

Method

Participants. Participants were 50 undergraduate students of the Erasmus University Rotterdam, The Netherlands (*M* age = 20.7, *SD* = 4.6). As in Experiment 1, participants had never played chess before. None of the participants took part in Experiment 1. The participants were randomly divided over two conditions.

Materials. The same materials were used as in Experiment 1.

Procedure. The procedure was similar to Experiment 1 with exception of the learning phase. The adjustments were made either to decrease interference between instructions, or to stimulate learners to generate cues that could aid in accurately monitoring understanding. First, participants were no longer required to provide self-explanations when providing move predictions. Although providing self-explanations when learning to play chess has been proven to have a positive effect on learning (De Bruin et al., in press), combining this instruction with the JOL and move selection instruction might have led to interference between the two tasks in Experiment 1. Moreover, the JOL and move selection instruction are assumed to have a similar effect on learning as self-explanation has, that is, to promote self-monitoring and self-regulation. Eliminating the self-explanation instruction allowed us to estimate the effect of providing JOLs and selecting moves for restudy without possible interference from the effect of self-explanation.

Secondly, before learners were required to provide JOLs, a generation phase was introduced that stimulated learners to generate cues about their level of understanding prior to monitoring comprehension in the JOL phase. Therefore, after predicting all moves for white in the prediction phase, and before the JOL phase, half of the participants were shown a mirror image of the same chess position and were once more asked to predict the move white had to make within 40 seconds (i.e., "generation task group"). The similarity of this task to the prediction phase ensured that the generated cues were relevant to monitoring comprehension. Moreover, the generation task immediately preceded the JOL instruction, so that the activated cues were still available at the time of providing JOLs. To ensure that participants self-generated these cues, they were not shown the actual moves of the computer. The other half of the participants underwent the same procedure, but did not generate predictions before providing comprehension ratings (i.e., "no-generation task group"). Instead, to equalize time on task between the groups, they were asked to count the number of legal moves the King and Queen could maximally make at every moment white had to move. After generating cues by predicting move 1 for white (generation task group) or after counting the number of legal moves white could make (no-generation task group), participants were asked to provide a JOL of move 1 for white. To maximize accuracy of self-regulation, participants were then immediately asked to indicate whether they wanted to study another example of move 1. Participants moved on to move 2, and had to either generate cues by predicting move 2 for white, or count the number of legal moves white could make, afterwards provide a JOL for move 2, and finally indicate whether they wanted to restudy move 2. This was repeated for all six moves of the endgame. The restudy of the selected moves was delayed until all move selections had been made. After restudying the selections, participants moved on to a new chess exercise, and the same procedure was repeated (first, predicting the moves of the complete endgame, then generating cues by predicting move 1, providing JOL move 1, selecting move 1 for restudy, etc.). As in Experiment

1, participants studied four chess exercises under these instructions, and afterwards played four versions of the endgame against the computer. Since the procedure in Experiment 2 required participants to estimate how well they had understood the move they just predicted, participants rated on a scale from 1 to 7 how well they had understood the predicted move (cf. Thiede et al., 2003, 1 = not at all understood to 7 = completely understood). An overview of the procedure of this Experiment is found in Table 1.

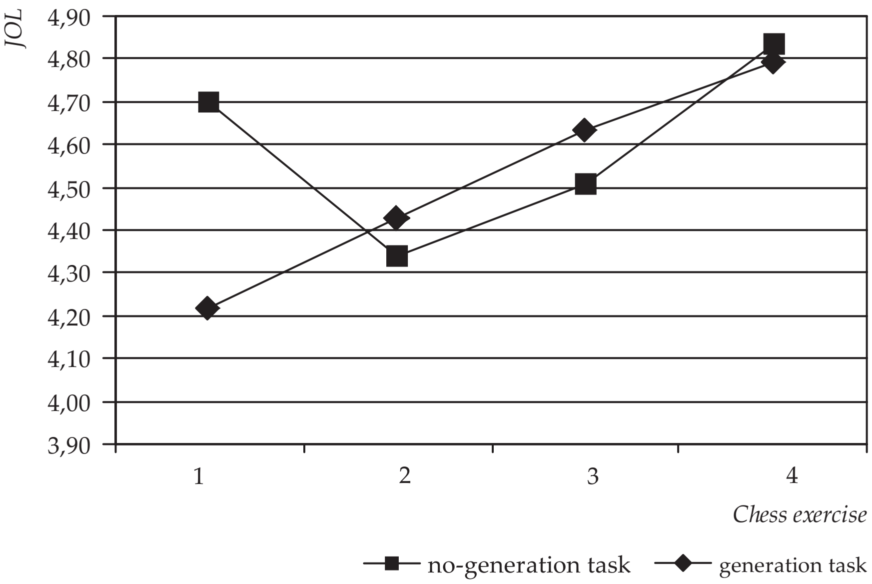
Analysis. The analysis of the move predictions was similar to Experiment 1. However, the repredictions were now predictions participants made in the generation task. The no-generation task group did not provide predictions in this phase. Since we could not compare the metacomprehension accuracy of the two groups, we compared the absolute JOL scores they produced. The remainder of the statistical analyses in Experiment 2 was similar to the analyses in Experiment 1, but with generation task instruction (yes or no) rather than expertise level as the between-subjects factor.

Results

Learning phase. Mastery of the basic chess rules did not differ between the two groups, F (1, 48) < 1. For one of the participants in the no-generation task group, results for the learning phase were not available due to technical problems. The analysis of the number of moves selected for restudy showed that the effect of the generation task was not significant, F (1, 47) = 2.22, MSE = 3.65, p = .14, η^2_p = .05. Thus, the group that had to generate predictions before providing JOLs did not select more moves for restudy than the group that did not generate predictions before JOLs. The effect of chess exercise was significant, F (1, 47) = 6.04, MSE = 1.91, p < .05, η^2_p = .11. As in Experiment 1, the number of move selections decreased over the four exercises. The analysis of absolute JOL scores showed no main effect of generation task, F (1, 47) < 1. However, an effect of chess exercise was revealed, F (1, 47) = 10.84, MSE = 0.35, p < .01, η^2_p = .19: the JOLs increased over the chess exercises. The interaction between group and chess exercise revealed a trend, F (1, 47) = 1.14, MSE = 0.35, p = .08, η^2_p = .07. A graphical presentation of these data is presented in Figure 2.

Figure 2

Absolute JOL scores in Experiment 2



As to the prediction scores in the learning phase, no main effect of generation task instruction was found, F (1, 47) < 1, but a significant effect of chess exercise was observed, F (1, 47) = 135.82, MSE = 310.65, $p < .001$, $\eta^2_p = .74$. That is, the prediction scores increased over the four exercises. The restudy scores revealed no main effect of the generation task, F (1, 22) = 1.19, MSE = 697.73, $p = .29$, $\eta^2_p = .05$. The effect of chess exercise was significant, F (1, 22) = 4.89, MSE = 405.41, $p < .05$, $\eta^2_p = .18$. However, since the latter analysis was based on the subset of participants who selected moves on all four exercises, results should be interpreted with caution. In Table 2, the mean move selections, JOLs, and prediction scores are presented for the two groups.

The monitoring accuracy and self-regulation correlations were calculated as in Experiment 1. Since the no-generation task group did not provide reprediction scores, metacomprehension accuracy correlations were not computed for this group. The correlations are shown in Table 4. The metacomprehension accuracy correlations for the generation task group are higher than those of the inexperienced chess players in Experiment 1 (mean r = .38 in Experiment 2 compared to r = .14 in Experiment 1), and seemed to increase over the four chess exercises. Despite the increased metacomprehension accuracy, the self-regulation correlations were still close to zero. Apparently, learners were better able to judge their comprehension level as a result of the generation task, but did not use that information to regulate study behavior. This is mostly observed in the low number of moves learners selected for restudy.

Table 4

Monitoring accuracy and self-regulation correlations for the Generation task group and the No-generation task group in Experiment 2. Monitoring accuracy is expressed through the correlation between JOLs and reprediction scores. Self-regulation is expressed through the correlation between JOLs and move selections. P values between parentheses

Expertise level	**Monitoring accuracy**			
	Exercise 1	*Exercise 2*	*Exercise 3*	*Exercise 4*
Generation task group	.39 (.06)	.27 (.20)	.44 (.03)	.42 (.04)
	Self-regulation			
	Exercise 1	*Exercise 2*	*Exercise 3*	*Exercise 4*
Generation task group	-.06 (.79)	-.20 (.34)	.07 (.76)	-.14 (.50)
No-generation task group	.36 (.09)	.18 (.40)	-.23 (.28)	-.30 (.15)

Test phase. A oneway analysis of variance showed that the generation task group more often checkmated the black King than the no-generation task group, F (1, 48) = 7.65, MSE = 0.85, p < .01, η^2_p = .14. On average, the generation task group checkmated the black King 1.60 times (SD = 1.00), and the no-generation task group 0.88 times (SD = 0.83).

Discussion

The results show that the instruction to generate a prediction before providing JOLs had a positive effect on learning to play an endgame of chess. That is, those who generated predictions before providing JOLs checkmated the black King more often in the transfer test than those who counted the number of moves white could make before providing JOLs. Moreover, the metacomprehension accuracy correlations of the generation task group were higher compared to those in Experiment 1. The absolute JOL scores of the generation task group showed a pattern that was more consistent with understanding than the JOL scores of the no-generation task group. For the former group, JOLs increased steadily with practice, whereas for the latter group this was not the case (see Figure 2). Apparently, activating cues that are predictive of future chess playing performance prior to rating comprehension improved metacomprehension accuracy. This, in turn, fostered understanding of the endgame. The question remains whether the higher test performance of the generation task group was due to the higher metacomprehension accuracy or to the act of predicting moves in the generation task. There are two arguments that support the first explanation. First, participants did not receive computer feedback on the predictions they generated prior to rating comprehension. Second, all other information the generation task group received was identical to the information the no-generation task group received. Therefore, the performance difference more than likely resulted from internal, metacognitive activity in the generation task group.

Although metacomprehension accuracy increased under the generation instruction, the correlations that indicated self-regulation did not improve. This was mainly due to the low number of moves participants selected for restudy. Even though participants apparently learned to estimate their level of understanding, they were unable to use this increased understanding to appropriately select moves that needed restudy. A possible explanation for this lack of self-regulation could be that learners were not aware of the accuracy of their JOLs. In studying translation pairs, for example, learners can readily determine whether their estimate of understanding is in line with current performance through attempting to retrieve the correct translation. If the learner produced a high JOL, but retrieval fails, he will realize that the JOL was inaccurate. This performance feedback will aid the learner to more correctly estimate understanding in future learning trials. In a cognitive skill such as chess, assessing the correctness of JOLs is more complex. First, the learner has to assess accuracy of the move prediction by analyzing the computer move. In order to improve metacomprehension accuracy, the learner should also evaluate the JOL accuracy by comparing the JOL to the quality of the prediction. Information for the first task is probably more salient, since the prediction of the move received more extensive processing attention than the production of a JOL. Possibly, for inexperienced chess players, this task poses high demands on processing resources, which disrupts execution of the second task.

One can deduce from this that, if processing requirements necessary to assess accuracy of the move predictions are reduced, resources will become available to assess accuracy of JOLs. The latter will positively affect metacomprehension accuracy, and will also lead to more accurate self-regulatory behavior. Therefore, in Experiment 3, the availability of performance feedback was manipulated. Half of the participants were provided with performance feedback after providing JOLs. That is, learners were not required to compare their predictions to the computer moves, but the computer assigned a quality score to the move predictions. The other half of the participants did not receive performance feedback. Providing feedback about the correct answer might reduce the need to self-generate move predictions. Therefore, we only assigned a quality score to the move predictions without giving the correct answer. The effect of the performance feedback was hypothesized to be twofold. First, it would prevent learners from entertaining illusions of understanding that might occur when they themselves assess the accuracy of their predictions. Therefore, the performance feedback would lead to better self-regulation (i.e., selecting those moves for restudy that learners predicted poorly). Moreover, not assessing the accuracy of the predictions frees up processing resources that can be allocated to assess the accuracy of the JOLs. This was hypothesized to improve metacomprehension accuracy. The close proximity in time of the performance feedback to the JOLs would enhance chances of comparison. In sum, the improved metacomprehension accuracy and better self-regulation would cause more thorough understanding of the KQK endgame and would, therefore, lead to better performance in the test phase.

Experiment 3

Method

Participants. Participants were 43 undergraduate students of the Erasmus University Rotterdam, The Netherlands (*M* age 19.8, *SD* = 1.6). As in Experiment 2, participants had never played chess before. None of these participants took part in Experiment 1 or 2. The participants were randomly divided over two conditions.

Materials. The same materials were used as in Experiment 1. Since the previous experiments showed that monitoring accuracy tended to increase over the learning cycles, an extra chess exercise of similar difficulty was added to the learning phase. This would allow participants to further improve monitoring accuracy and self-regulatory activities.

Procedure. As in Experiments 1 and 2, participants started by studying the basic chess rules, and were afterwards tested on their knowledge of these rules. The procedure of the learning phase of Experiment 3 differed in one respect from the procedure of the generation task group in Experiment 2. To minimize illusions of understanding, and to free up processing resources that could be allocated to evaluate the accuracy of JOLs, half of the learners were provided with performance feedback about move predictions. Thus, in the feedback group, after learners had predicted a move and provided a JOL for that move in the generation phase, the computer provided feedback about the quality of the move prediction by assigning a score between one and four. A score of one indicated an illegal move, or an incorrectly typed answer. A score of two indicated a bad move, corresponding with a move that led to a situation that needed two or more extra moves above the minimum needed to checkmate. A score of three was labelled a suboptimal move and represented a move that led to a situation that needed one extra move than the minimum to checkmate. Finally, a score of four indicated the best move possible. Note that this scoring procedure corresponds to the procedure used in all experiments to analyze move predictions. At the top of the screen, the score was projected, and at the bottom the labelling of the scores was explained. Now, illusions of understanding were prevented, and participants were stimulated to compare the recently provided JOL to the performance feedback. We hypothesized that this would increase the accuracy of JOLs over the course of the learning phase. Moreover, the feedback would stimulate self-regulation by signalling to learners when they needed to select a move for restudy. The other half of the learners followed the same procedure, but did not receive performance feedback (i.e., no-feedback group).

Analysis. The analysis of the move predictions as well as the statistical analyses were similar to Experiments 1 and 2.

Results

Learning phase. The same analyses were performed as in the second experiment, with feedback (yes or no) rather than generation task as the between-subjects factor. On the basic chess rules test, the no-feedback group performed significantly higher than the feedback group, F (1, 41) = 4.30, MSE = 0.88, $p < .05$, $\eta^2_p = .10$. The no-feedback group had a mean of 8.55 on a scale from 1 to 9 (SD = 0.60), whereas the feedback group had a mean of 7.96 (SD = 1.15). Since any errors made on the basic chess rules test were immediately explained by the experimenter before proceeding to the learning phase, this difference probably did not affect the learning phase.

The analysis of the number of moves selected for restudy showed that the effect of feedback was not significant, F (1, 41) = 103.04, MSE = 7.62, $p = .31$, $\eta^2_p = .03$. There were no differences in number of move selections between those who received feedback and those who did not. However, the effect of chess exercise was significant, F (1, 41) = 33.43, MSE = 1.78, $p < .001$, $\eta^2_p = .45$. As in the first two experiments, the number of move selections decreased over the exercises.

A trend was observed for the main effect of feedback for the prediction scores in the learning phase, F (1, 41) = 2.89, MSE = 760.24, $p = .10$, $\eta^2_p = .07$. A significant effect of chess exercise was also revealed, F (1, 41) = 62.32, MSE = 373.86, $p < .001$, $\eta^2_p = .60$. Due to technical problems, data were missing for the reprediction scores for three of the participants in the no-feedback group. The reprediction scores again revealed a trend for the effect of feedback, F (1, 38) = 2.96, MSE = 1322.76, p = .09, η^2_p = .07. The reprediction scores increased significantly over the course of the chess exercises, F (1, 38) = 61.70, MSE = 222.48, p < .001, η^2_p = .62.

For the restudy scores no main effect of feedback was observed, F (1, 22) < 1. However, these data showed a quadratic effect of chess exercise, F (1, 22) = 52.55, MSE = 389.85, p < .001, η^2_p = .71. That is, the restudy scores increased over the first four chess exercises, but decreased for the final chess exercise. However, since this analysis was based on the subset of participants who selected moves on all four exercises, results should be interpreted with caution. The mean scores for these variables are shown in Table 2.

We calculated the metacomprehension accuracy and self-regulation correlations as in Experiments 1 and 2. These correlations are shown in Table 5. As to metacomprehension accuracy, there was no difference between the two groups (mean r = 0.28 for the feedback group, mean r = .29 for the no-feedback group). The self-regulation correlations indicate an advantage for the no-feedback group (mean r = -.03 for the feedback group, mean r = -.31 for the no-feedback group). Also, for the no-feedback group, the correlations increase somewhat over the course of the exercises. In Figure 3, the mean metacognitive correlations (metacomprehension accuracy and self-regulation) are presented graphically for the three experiments.

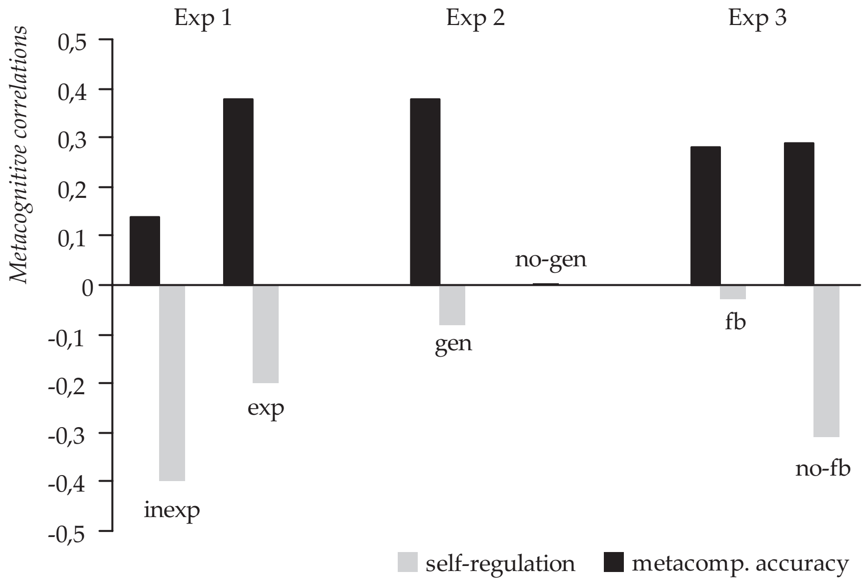
Table 5

Monitoring accuracy and self-regulation correlations for the Feedback group and the No-feedback group in Experiment 3. Monitoring accuracy is expressed through the correlation between JOLs and reprediction scores. Self-regulation is expressed through the correlation between JOLs and move selections. P values between parentheses

Condition	**Monitoring accuracy**				
	Exercise 1	*Exercise 2*	*Exercise 3*	*Exercise 4*	*Exercise 5*
Feedback group	-.01	.30	.38	.35	.36
	(.98)	(.18)	(.07)	(.10)	(.09)
No-feedback group	.28	.39	-.03	.23	.59
	(.24)	(.10)	(.90)	(.35)	(.01)
	Self-regulation				
	Exercise 1	*Exercise 2*	*Exercise 3*	*Exercise 4*	*Exercise 5*
Feedback group	.01	.12	.01	-.18	-.08
	(.96)	(.60)	(.96)	(.40)	(.71)
No-feedback group	-.13	-.05	-.43	-.57	-.40
	(.60)	(.83)	(.06)	(.01)	(.09)

Figure 3

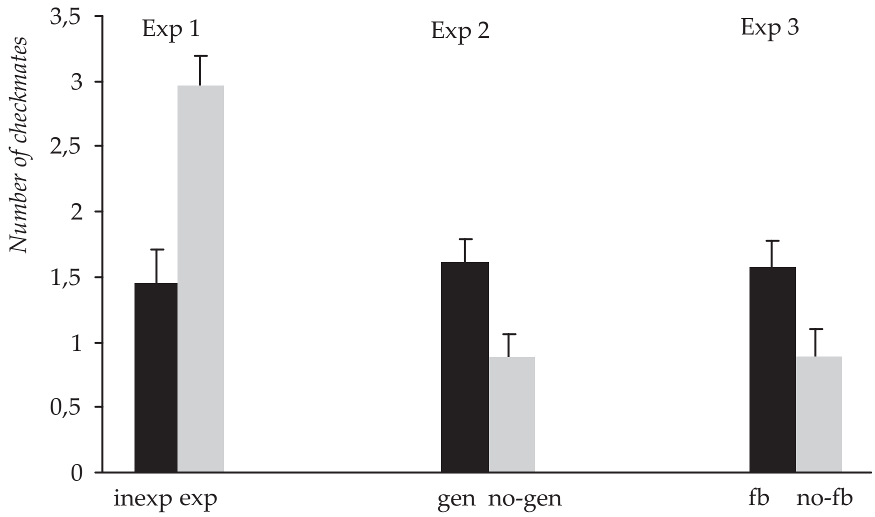
Mean metacognitive correlations for the three experiments. Note: inexp = inexperienced chess players, exp = experienced chess players, gen = generation task group, no-gen = No Generation task group, fb = Feedback group, no-fb = No Feedback group



Test phase. A oneway analysis of variance showed that the feedback group more often checkmated the black King than the no-feedback group, F (1, 41) = 6.61, MSE = 0.95, $p < .05$, $\eta^2_p = .13$. The mean number of checkmates for the feedback group was 1.57 (SD = 1.16), and 0.80 (SD = 0.70) for the no-feedback group. In Figure 4, the test results for the three experiments are represented.

Figure 4

Number of checkmates in the test phase for the three experiments. Note: inexp = inexperienced chess players, exp = experienced chess players, gen = Generation task group, no-gen = No-generation task group, fb = Feedback group, no-fb = No-feedback group



Discussion

In line with our expectations, the feedback instruction positively affected test performance in novices studying a chess endgame. Possibly, the performance feedback helped learners to evaluate the correctness of their predictions, and reject or confirm hypotheses about underlying chess principles. The difference in learning rate as indicated by the prediction scores in the learning phase was small, but showed a trend. Despite the positive effect of feedback on transfer performance, it did not improve metacomprehension accuracy or self-regulation over the control instruction. That is, there were no differences in metacomprehension accuracy between those who did and those who did not receive feedback, and the group that did not receive feedback even had higher self-regulation correlations than the feedback group. The latter finding could not be attributed to a higher number of restudies in the no feedback group, since these did not differ between the groups. A possible explanation could be that the performance feedback impeded the development of self-evaluative skills, because the evaluation was performed by the computer (Mathan & Koedinger, 2005). That is, learners were not stimulated to self-evaluate the accuracy of their predictions during the JOL phase, but awaited the computer feedback to do it for them. Instead, the performance feedback might have promoted the development of generative skills (Mathan & Koedinger, 2005). Generative skills are skill components directed at selecting appropriate problem solving operators in specific task situations. The generative skills assumption is substantiated by the finding that learners who received performance feedback were better able to transfer the acquired principles to the test exercises. The processing resources that became available as a result of feedback were possibly not allocated to developing metacomprehension skills, but to developing knowledge of the chess principles. In sum, this study suggests that performance feedback positively affects transfer performance, but does not improve metacognitive activities as metacomprehension and self-regulation over the generation task instruction tested in Experiment 2.

General Discussion

The experiments undertaken in this study were designed to examine novices' ability to benefit from metacognitive strategy training in acquiring a cognitive skill as chess. Analyzing possible a priori differences in use of metacognitive strategies between novices and more experienced chess players revealed that the experienced group showed higher metacognitive awareness, but not better self-regulation of learning activities. Furthermore, the metacomprehension accuracy of the experienced chess players showed room for improvement. Apparently, other factors besides having proceduralized the rules of the domain determine onset of effective use of metacognitive strategies in chess skill acquisition. More specific instructions are needed that prompt learners to monitor and regulate learning activities.

Based on the cue-utilization framework (Koriat, 1997), Experiments 2 and 3 examined the effect of specific instructions designed to improve metacognitive activities in chess skill acquisition. Experiment 2 revealed that when generating a move prediction prior to rating comprehension, learners not only more accurately monitored comprehension of the moves, but also performed better in the test that required them to transfer the inferred chess principles to novel games. Possibly, the generation instruction activated relevant cues that indicated how well the move was predicted, and, therefore, how well it would be predicted on a future trial. For example, when learners had little difficulty predicting a move, the cues indicated that the problem was solved easily. This ease-of-problem-solving judgment was still available when assessing comprehension of the predicted move immediately after the generation task, and, thereby, improved metacomprehension accuracy. The essence of self-generating these cues is emphasized in this line of reasoning: For the development of metacognitive skills, it is crucial that learners themselves practice determining how easy or difficult solving the prediction problem was. This is further substantiated by the lack of an effect of performance feedback on metacomprehension accuracy in the third experiment: When cues were generated externally by the computer, learners were not stimulated to use them when judging comprehension, and metacomprehension accuracy did not improve. The performance feedback informed learners only about the result of their problem solving and not about the efficiency or invested effort. A mismatch between the ease-of-processing cues learners activated and the quality-of-processing cues the performance feedback generated might have been at stake here.

The lower metacomprehension correlations relative to those in memory or text processing research are not entirely surprising in light of the lack of experience of our participants, and the poverty of the mnemonic cues that they could activate. From the perspective of the cue-utilization framework, diverse cues as experienced familiarity, or memory for outcome of previous recall attempts may simultaneously be used to judge comprehension (Koriat, 1997). In the present experiment, learners could not rely on cues from previous experiences, but only on ease-of-problem-solving cues from the current learning experience. Hence, metacomprehension accuracy, although improving under the generation task instruction, was lower compared to previous studies were multiple cues were available to learners (e.g., Thiede et al., 2003). Note that, despite the relatively lower correlations, higher metacomprehension accuracy was related to higher transfer performance.

Despite the improvement in metacomprehension accuracy and transfer performance, the generation instruction left the self-regulation activities of the learners unaffected. Only in the third experiment, we saw an increase in self-regulation correlations in the group that did not receive feedback. However, the lower test performance of this group compared to the feedback group calls into question the effect of self-regulation on chess skill acquisition. A possible explanation for this disappointing result may lie in the inappropriateness of the restudy activities. In memorizing word pairs or studying expository text, restudy of the material directly leads to resolving uncertainties in knowledge, and, therefore, has a clear relation to performance improvement. In chess skill acquisition, and probably in other domains of cognitive skill acquisition, the relation between restudy and performance is less clear-cut. Learners had to infer the chess principles by studying the computer chess moves. Even though learners judged that a particular move was not understood, they may also have realized that simply restudying the computer move would not supply the information needed to improve understanding. As a result, learners refrained from restudy of the moves. Instead, if restudy provided more explicit explanation of the principles, learners might have selected more and more adequately. Previous studies assumed that adequate self-regulation would logically result from accurate metacomprehension. The present study shows that, in chess skill acquisition, this relation not necessarily exists. Future research should test the inappropriateness hypothesis by redesigning the restudy phase to adapt it more to learners' restudy needs.

The absence of an effect of feedback on metacognitive activities, combined with a positive influence of feedback on transfer performance calls into question the applicability of metacognitive strategy instruction in low-knowledge learners acquiring a cognitive skill. Several authors have questioned the use of pure discovery learning in low knowledge students, and argue for direct instruction (Klahr & Nigam, 2004) or guided discovery (Mayer, 2004). In their view, a lack of external guidance entails the risk of not actively processing the information, and, thereby, failing to discover the principles. Therefore, a more structured learning situation is needed that ensures learning of the domain principles. Even though the instructions applied in the present experiments contain elements of discovery learning, they represent a guided discovery approach: Learners are provided feedback about their predictions, and are stimulated to improve metacognitive awareness and self-regulate to aid discovery of the chess principles through the generation task. Moreover, these instructions appear effective: Improving metacomprehension in low prior knowledge learners on a complex task enhances transfer performance. In sum, when fostering chess skill acquisition, directing instruction at improving metacognitive awareness seems worthwhile.

The fact that the feedback instruction did not affect metacomprehension, but did improve transfer performance raises doubt about the long-term retention of the inferred principles under feedback instructions. The feedback might have stimulated superficial processing during learning, which benefited short-term transfer performance, but might have impeded long-term retention (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Bjork, 1994; Rosenbaum, Carlson, & Gilmore, 2000). To examine this possibility, a combination of short-term and long-term transfer tasks in future research is needed. This will also control for any effects due to diminished concentration on the short-term transfer test. For example, in Experiment 3, the control group scored relatively lower than the comparable generation task group in Experiment 2 on the transfer test. The only difference that existed between these groups and that could have caused this result is mental fatigue due to an extra chess exercise in the learning phase that affected performance on the test phase.

In the current experiments, for the first time, the applicability of the cue-utilization framework (Koriat, 1997) is demonstrated for cognitive skill acquisition. Moreover, the generation task instruction not only provided insight into the cognitive mechanisms that mediate metacognitive monitoring, but also proposed a design principle that can be easily transferred to other domains of metacognitive training. That is, even for novices in a domain, the development of metacognitive monitoring in cognitive skill acquisition can be improved by requiring learners to generate relevant cues about the test task immediately prior to providing comprehension ratings. This will not only enhance learning, but will also increase chances of transfer to novel situations. Future research should test the applicability of the cue-utilization framework in other skill domains besides chess, and, given the ultimate goal of transfer of learned skills to future situations, should also take into account long-term tests of performance.

Chapter 7

Summary and general discussion

The studies reported in this dissertation were undertaken to shed light on some important issues that currently prevail in research on the development of expertise, and in research on fostering metacognitive processes in learners in particular. In general, the studies reported here were developed to determine to what extent, and in which format, novice learners can benefit from engaging in metacognitive processes as employed by experts. Departing from principles on learning and instruction as postulated in the theory of deliberate practice (Ericsson et al., 1993), our goal was to formulate and test implementations of this theory in non-expert learners. Based on an analysis of experts' practice behavior, Ericsson and colleagues postulated that experts' exceptional performance is not explained by a genetic predisposition, but can rather be ascribed to years of dedication to intense practice schedules. Particularly when practice is deliberate and centered around tasks that are designed to improve performance (compared to tasks performed for fun), the development of expertise will be maximized. More specifically, when the study task is of an appropriate difficulty level, when the learner receives informative feedback about performance, and when there is ample room for repetition and correction of errors, this type of practice is referred to as "deliberate practice". Ericsson and colleagues (1993) showed that the total number of hours spent on deliberate practice differentiates individuals of varying levels of expertise. That is, a monotonic positive relationship was observed between amount of deliberate practice and performance level. This monotonic benefits assumption has been tested and confirmed in a variety of studies in diverse settings (e.g., Charness et al., 1996, 2005; Helsen et al., 1998; Hodge & Deakin, 1998; Hodges et al., 2004).

As a rather extensive base of evidence has been established for the relation between investments in deliberate practice and experts' improvements in performance, a next step in research would be to determine to what extent this association is also observed in non-experts. Drawing this line even further, it would be of interest to education, to study under what circumstances novices can benefit from learning strategies that are initially considered characteristic of experts. Although theoreticians have mainly questioned novices' abilities to acquire metacognitive skills for a number of reasons (e.g., Kalyuga et al., 2003; Kanfer & Ackerman, 1989; Winne, 1995a), recent studies paint a somewhat more optimistic picture, and conclude that, under certain conditions, low-knowledge learners are able to successfully apply metacognitive strategies in order to maximize their learning results (e.g., Azevedo, 2004; Butler, 1998; Mathan & Koedinger, 2005). That is, if novices' lack of hierarchically organized knowledge schemes and, therefore, easily overloaded working memory are taken into account, and if specific principles from research on metacognition are applied (e.g., Koriat, 1997), the result will be a tailored instruction that will potentially stimulate metacognitive processes in low-prior knowledge learners. In general, a scientific test of such an instruction will provide insight into how, and to what extent, expert-learning strategies can be fostered in novices.

Overview of the main results

Prior to testing the effect of fostering metacognitive processes in chess novices on learning and transfer performance, it is worthwhile to examine in a real-life field setting whether non-experts are able to benefit from expert-like learning strategies, such as deliberate practice. To that end, the study in **Chapter 2** analyzed, by means of a linear mixed models analysis, the relation between elite, adolescent chess players' investments in deliberate practice and chess performance ratings, while taking into account their development over time. This analysis allowed us to determine the contribution of deliberate practice, not only to current level of performance, but also to chess performance at time points early on in their career. Moreover, this study was undertaken to test one of the main critiques raised against the assumptions of the deliberate practice theory, which states that the factors talent and deliberate practice might have been confounded in previous research (Sternberg, 1996). By relying solely on information from those who have succeeded and attained expert level, the theory of deliberate practice might be based on a subset of the population that could be characterized by a predisposition to benefit from investments in deliberate practice. This hypothesis can be tested by comparing the effect of similar investments in deliberate practice on performance for those who have achieved expert performance to those who have given up at some point. In Chapter 2, this comparison was made between a group of elite adolescent chess players who still attended the highest level of Dutch national chess training and a comparable group of chess players who had voluntarily dropped out of this training.

The findings revealed that throughout chess players' careers, deliberate practice proved to reliably predict chess performance measured by participants' chess ratings. A model that included investments in deliberate practice and their fluctuation over time explained more than 70% of the variance in chess ratings. Previous studies that did not include the time factor usually accounted for less variance; maximally up to 60% (e.g., Charness et al., 1996, 2005; Hodges et al., 2004). These results imply that even at lower levels of performance (in this study, at earlier time periods in career) deliberate practice reliably predicted chess performance. Moreover, for both persistent and dropout chess players, equal investments in deliberate practice had a comparable effect on performance. Rather than differing in how much they benefit *from* deliberate practice, these groups were distinguished by total time invested *in* deliberate practice. These findings are not in line with Sternberg's assumption (1996) that experts might be characterized by a talent to maximally benefit from training. Finally, we examined in Chapter 2 the possible contribution of deliberate practice to performance differences typically detected between males and females in chess. Our analysis confirmed that male chess players generally performed higher than female chess players, but also that this difference was mainly explained by variation in dedication to deliberate practice between males and females: The results revealed that males tended to spend more time practicing deliberately than females.

Previous research has largely disregarded the potentially crucial role motivation plays in maintaining high levels of deliberate practice. Therefore, the study in **Chapter 3** investigated the influence of general achievement motivation and motivation to specifically practice chess on accumulated hours of deliberate practice, and current performance level in a group of elite adolescent chess players. Although experimental research raises questions as to the effect of competition on intrinsic motivation, and hence, time involvement in a task (e.g., Deci et al., 1981; Reeve et al., 1985; Vallerand & Reid, 1984), applied studies in sports suggest that competitiveness might be characteristic of individuals performing at exceptional levels (e.g., Brewer et al., 1993; Gould et al., 2002). To further address this issue, the study in Chapter 3 also examined to what extent a competitive achievement orientation contributed to investments in deliberate practice, by means of structural equation modeling. Our findings indicated that achievement motivation distinguished between persistent and dropout chess players, and moreover, that both chess-specific motivation and achievement motivation strongly influenced investments in deliberate practice. A model that depicted motivation as the result, instead of the cause, of deliberate practice and performance provided a lower fit to the data. This study provided information on some of the necessary conditions for deliberate practice that deserve attention in training and education. Given that our results underline that when motivation diminishes, engagement in deliberate practice decreases, this leads to a recommendation for trainers and teachers to frequently monitor motivation in their students.

Having established that deliberate practice also proves beneficial to performance improvement in individuals of lower performance levels, we set out in the last three studies to explore the effect of fostering metacognitive processes in chess novices on learning and transfer performance. In **Chapter 4**, novice chess players were required to study various computer played examples of the chess endgame of King and Rook against King under one of three experimental procedures. The first group predicted the moves of the computer while self-explaining their predictions. That is, they explained aloud to themselves why they believed the computer would make a certain move. The second group only predicted the moves of the computer without self-explaining, whereas the third group merely observed the moves of the computer without predicting or self-explaining. We hypothesized that the group that self-explained and predicted the moves of the computer would outperform the group that merely predicted, because they were stimulated more to explicitly reflect on the explanations for their predictions and on the moves of the computer. The latter group would, in turn, perform higher than the group that only observed the moves of the computer, because the group that only observed the moves was not explicitly required to perform any kind of problem solving behavior. Our results indicated that, compared to the two other groups, those who self-explained and predicted not only inferred the principles of the chess endgame more readily in the learning phase, but also checkmated the black King more often in the transfer test. No differences were observed between the two remaining groups. These findings suggest that despite the demand it might pose on working memory, self-explaining while acquiring a complex cognitive skill as playing chess stimulates learning of domain principles. Merely predicting without self-explaining possibly does not invoke the amount of task involvement in learners that is needed to evaluate the accuracy of the inferred principles. For example, when asked to predict moves without verbalizing explanations for the predictions, learners' self-monitoring remained at a more superficial level, resulting in learning gains comparable to those who merely observed the moves of the computer.

Whereas the study in Chapter 4 provided insight into novices' capacities to optimize learning gains by fostering learners' self-awareness in general, the studies described in Chapter 5 and 6 were set up to stimulate development of specific metacognitive activities in novices. On the basis of previous studies conducted in the domain of memorizing word pairs (e.g., Son & Metcalfe, 2000; Thiede 1999; Thiede & Dunlosky, 1999), self-monitoring and self-regulation were identified as essential metacognitive skills that have a well-documented effect on performance. In the study described in **Chapter 5**, self-monitoring was operationalized by requiring novice chess players, after studying an example of the endgame of King and Queen against King, to first estimate how confident they were that they would correctly predict a similar move in the future. Next, this estimate (referred to as a "Judgment of Learning", JOL) was correlated with actual performance. Accurate monitoring was indicated by a high and positive correlation. Self-regulation was calculated by asking learners, after providing JOLs, whether they wished to study another example of the moves (yes = 1, no = 0), and correlating that answer with the comprehension ratings (JOLs). A high and negative correlation indicated adequate self-regulation; Learners chose those moves for restudy that they believed to have understood poorly.

The study reported in Chapter 5 contained two independent variables, each with two levels, resulting in a total of four groups of participants, who were all novices in chess. The first independent variable determined the effect of providing JOLs on metacognitive awareness and, hence, on learning. After studying an example of the chess endgame of King and Queen against King, two groups were asked to estimate for every move white had to make how confident they were that they would correctly predict a similar move in the future. To calculate a measure of monitoring accuracy, participants afterwards predicted the moves, which were scored and correlated with the JOLs. The two other groups did not provide JOLs. We hypothesized that those who were asked to provide JOLs would attain higher metacognitive awareness of their understanding of the moves, and would therefore better self-regulate their learning endeavors. This would ultimately result in faster and more thorough learning of the chess principles. The second independent factor determined the effect of forced versus free selection of moves for restudy. Given their limited experience with chess, the learners might have difficulty estimating to what extent they had understood the moves and how many needed restudying. Therefore, two groups were instructed to restudy at least two moves for every endgame of six moves. The two remaining groups were not instructed on the minimum number of move selections, but were free as to the number of moves they wanted to restudy. We hypothesized that those who were forced to select at least two moves would better regulate their restudies, by more accurately selecting the moves that indeed needed restudying, because they were stimulated more to reflect when selecting moves than the free selection groups.

Our results revealed that providing JOLs had a negative effect on predicting the correct moves in the learning phase, whereas being forced to select at least two moves for restudy improved the quality of the predictions over the course of the learning phase, even after controlling for actual number of restudies. In the transfer test, the forced selection groups outperformed the free selection groups by more often checkmating the black King. No differences were observed in the test phase for the JOL and No JOL groups. Moreover, the forced selection groups demonstrated higher monitoring accuracy and better self-regulation than the JOL groups. Apparently, for these groups of low prior-knowledge learners, providing JOLs did not improve metacognitive awareness, and to a certain extent even hampered learning of the endgame principles. By contrast, forcing low knowledge learners to select a minimum number of moves for restudy positively affected their self-reflection and self-regulation, which led to larger learning gains. These findings are discussed within the framework of the expertise reversal effect (Kalyuga et al., 2003). That is, transfer of expert learning strategies to novices and vice versa does not necessarily eventuate optimal learning because of differences in knowledge organization between experienced and inexperienced learners. Therefore, adaptation of expert instruction to novices' knowledge structure is required to maximize learning in inexperienced individuals.

To further examine possibilities of fostering expert-like metacognitive processes in novices, the study reported in **Chapter 6** describes the effect of an adapted version of the experimental procedure tested in Chapter 5 on novices learning to play chess. However, before examining the effect of this adaptation, the first experiment compared the monitoring accuracy and self-regulatory behavior of inexperienced and more experienced chess players when studying the endgame of King and Queen against King, to assess to what extent the effective use of these metacognitive skills depends on prior experience. Our findings demonstrated that, even for experienced chess players, correct use of these metacognitive skills is not straightforward and necessitates explicit practice. Therefore, in the second experiment we tested an adapted version of the procedure designed to stimulate metacognitive skills that was developed in Chapter 5. Based on the cue-utilization framework of metacognitive monitoring (Koriat, 1997), learners were required to activate relevant cues about their level of understanding immediately prior to rating comprehension of a particular move (i.e., prior to providing a JOL). That is, learners were once more asked to predict the move of the computer, but were not given feedback about the correct move. Since immediately afterwards they provided an estimate of how well they had understood the move, cues that were activated during prediction of the move and that indicated understanding were still active at the time of providing the JOL. In order to decrease working memory load, the move selection phase was placed directly after the JOL phase. Reactivation of relevant cues at the time of JOL or move selection was now no longer needed, as these were probably still available in working memory. Although the insertion of this so-called "generation task" prior to providing JOLs led to an improvement of monitoring accuracy (hereafter referred to as "metacomprehension accuracy") and higher performance on the test phase, self-regulation (i.e., the correct selection of moves that needed restudying) was still poor.

In the third and final experiment of Chapter 6, we set out to further diminish working memory load associated with the metacognitive tasks that were performed during study of the chess endgame. In experiments 1 and 2, the learning environment we provided was of a complex nature: Learners not only had to analyze to what extent their predictions corresponded with the computer's actual move, but were also required to deduce the correctness of their JOLs by comparing it to the computer moves. The duplicity and, hence, complexity of this task possibly disrupted execution of the latter task. Therefore, we provided learners with direct feedback on the accuracy of their predictions, so that they could direct processing resources to evaluating the correctness of their JOLs. We assigned a performance score (between 1 and 4, 1 = bad move, 4 = best move possible) to every move prediction. Given that this performance feedback was provided immediately after learners' had produced a JOL, conditions for comparison of feedback with JOL were optimized. We hypothesized that the performance feedback would free up processing resources to evaluate correctness of JOLs, and would thereby foster development of metacognitive monitoring and self-regulatory behavior. Contrary to our expectations, our findings showed that the effect of performance feedback on metacomprehension was limited, and, moreover, that self-regulatory skills lagged behind the group that was not given feedback. A recent review by Mathan and Koedinger (2005) suggests that immediate feedback might depress development of self-evaluative skills, because learners are no longer required to reflect on learning, but instead come to depend on external feedback. However, the feedback did show a positive influence on performance on the transfer test. This indicates that the feedback, despite the limited effect on metacognitive skills, fostered understanding of the chess endgame principles.

Discussion

Much has been said in the literature about the defining features and associated learning gains of self-regulated learners (Butler & Winne, 1995). These observations are, in our opinion, primarily relevant from the perspective of fostering these features in those who do not self-regulate their study behavior. Stated otherwise, insight into the metacognitive strategies expert learners apply during problem solving can serve as input for research on how to stimulate these processes in non-self-regulated learners. A typical example of such a cycle is observed in early research on the self-explanation effect. First, individual differences in spontaneous self-explaining were analyzed and related to learning gains (Chi et al., 1989). Having determined the positive effect of spontaneous self-explaining on learning, the study that followed assessed the effect of instructing individuals to self-explain while studying a text about the circulatory system (Chi et al., 1994). Given the positive effect of this technique on acquiring a mental model of the circulatory system, this study led to fruitful implications for further research and education. The studies reported in this dissertation follow a similar empirical cycle. First, the effect of expert learning strategies (i.e., deliberate practice) is determined in Chapters 2 and 3. In the remaining three studies in Chapters 4 through 6, the effectiveness of enhancing expert-like metacognitive processes is tested in several groups of novice chess players. This research cycle first establishes an empirical basis for the effect of a particular technique by analyzing its spontaneous use by certain individuals. Next, when the positive learning gains of the technique have been demonstrated, possibilities for instruction are explored in those who do not spontaneously apply it. The results of these studies allow for the formulation of both theoretical and educational implications. For the studies reported in this dissertation these implications are described below.

Although not an explicit goal of the research reported here, our studies in some ways support an environmentalist position within the nature-nurture debate on explanations for exceptional performance (for a review, see Howe, Davidson, & Sloboda, 1998). For instance, the studies in Chapters 2 and 3 underline the need for extensive domain-specific experience in order to achieve expert performance in chess. Chess-specific experience, especially of the kind that is termed "deliberate practice" proved to reliably predict variation in performance of elite, adolescent chess players. The factors gender and persistence contributed only marginally to this relation. Although chess-specific motivation and achievement motivation were found to have a relatively large influence on deliberate practice, there are no reasons to assume that these types of motivation are innate, and not influenced by environmental factors. Our results add to the long list of studies reporting extraordinary performance of individuals after extended domain-specific practice (e.g., Ericsson, Chase, & Faloon, 1980; Ericsson & Polson, 1988; Gordon, Valentine, & Wilding, 1984; Howe, Davidson, & Sloboda, 1998). These studies, as well as the studies reported in Chapters 2 and 3 of this dissertation, implicitly lead to the conclusion that variation in performance as the result of practice exceeds variation caused by individual differences. That is, although individual differences in abilities may exist, these are far less a determinant of performance than differences in investments in deliberate practice. The studies reported in Chapters 4, 5, and 6 further specify the domain-specificity assumption in two respects. First, these studies demonstrate that the relation between domain-specific experience and performance also holds for individuals of low prior knowledge. That is, after intense practice designed to foster expertise development, absolute novices in chess proved to be able to deduce and correctly utilize the domain principles of a chess endgame.

Moreover, our studies indicate that the domain-specificity principle also applies to the development of metacognitive skills as self-monitoring and self-regulation. Through domain experience, the novice chess players' metacognitive processes improved, and enabled them to transfer the chess principles to novel problems. Furthermore, apart from the positive effect on comprehension of the endgame principles, repeated cycles of self-monitoring and self-regulation seemed to enhance the accuracy of these metacognitive processes. All in all, this leads us to conclude that the skills of the elite group of learners referred to as self-regulated learners (e.g., Butler & Winne, 1995) need not be described as being innate or resistant to instruction. Our studies underline that, through proper training, these skills can be taught to non-self-regulated learners. However, as the spontaneous use of these skills is low and transfer to novel situations is rare, extensive domain experience is required in order to develop these skills adequately. Below, we will describe a number of conclusions that resulted from our studies and that are relevant when designing metacognitive strategy training for novices. These guidelines are largely based on the results of the studies reported in this dissertation, supplemented with insights from previous research.

Conclusions for metacognitive strategy instruction in novices

First, to optimize chances of implementation in education and to decrease external regulation, the instruction should be as teacher-independent as possible. That is, the learner should, after a brief instructional session, be able to apply the strategies by him/herself, without the need for guidance by a tutor. To ensure thorough understanding of the strategy and to improve chances of transfer, the learner should self-direct learning of the metacognitive skills. Moreover, the teacher-independence will increase chances of educational implementation, as it will not burden teachers' already heavily loaded work schedule. Also, the teacher-independence will induce instructional simplicity that allows for application of the strategy in other domains. Because of its low domain-dependence, the strategy training, as tested in chess, will probably prove beneficial to learning in other cognitive skill domains after simple adaptations. In the present studies, the teacher-independence was mainly obvious in Chapter 4, where learners self-explained while predicting computer chess moves. Requiring learners to effectively self-explain necessitates an instruction by a teacher, preferably including an example of a proficient model (Bielaczyc et al., 1995), but all in all this instruction need not take longer than approximately ten minutes. Although teacher-independent, the procedures intended to foster metacognitive processes in novices as tested in Chapters 5 and 6 are not directly transferable to a self-study situation because of its dependence on a computer. The participants worked through a computer-based chess application, and received all relevant information on the screen. However, with minor adaptations and a minimum amount of instruction, the crucial aspects of the procedure in Chapters 5 and 6 (i.e., stimulating self-monitoring and self-regulation of study through providing JOLs and selecting moves for restudy) should be transferable to computer independent self-study. For example, if learners are taught to divide study material into meaningful units (i.e., paragraphs of a text, or separate exercises in math or physics) and self-rate comprehension of these units, they can also be taught to determine, on the basis of the comprehension ratings, which parts of the material need restudying. Thus, even though the procedures tested in Chapters 5 and 6 were implemented in a computer-based environment, these can fairly easily be translated to a computer-free self-study situation.

Given the absence of automated problem solving procedures that characterizes novice learners, any instruction intended to increase learners' metacognitive awareness above their regular problem solving endeavors should place as low a demand as possible on individuals' working memory. Any unnecessary extra demands will undoubtedly hamper learning (Sweller, 1999; Sweller, Van Merriënboer, & Paas, 1998). We observed this effect in Chapter 5, where novice chess players' learning was seen to suffer from the requirement to monitor understanding by providing JOLs. Only when specific precautions were taken to minimize working memory load in Chapter 6, the effect of the metacognitive strategy training appeared.

The fact that providing self-explanations as tested in Chapter 4 proved to positively affect learning might be considered contradictory to this conclusion at first sight, given that verbalizing thoughts while solving a problem in an unknown domain will likely pose a heavy demand on working memory. However, unlike the metacognitive activities required from learners in Chapters 5 and 6, the act of self-explaining in Chapter 4 was entirely self-directed. That is, learners were instructed to self-explain continuously, but the content and the extensiveness of the self-explanations were determined by learners themselves, depending on their available processing resources. If learners had difficulty with the task, little working memory capacity was available for self-explaining, probably causing the number of self-explanations to drop. The high level of self-directedness and the accompanying adaptation to individuals' level of learning prevent working memory overload and explain large part of the effectiveness of self-explanations above explanations generated by a teacher (for an overview of the advantages and disadvantages of self-explanations over instructional explanations, see Renkl, 2002).

Related to the working memory issue, it is imperative that the metacognitive strategies can be automated through repeated practice in order to be useful.

As mentioned by Boekaerts (1997), a large proportion of expert learners' metacognitive strategies are not deliberate, but automated and habitual. That is what accounts for part of their effect in demanding and complex learning environments. Therefore, for metacognitive strategy training to have an effect on learning, not only in the current learning environment, but also in future situations, opportunities for automation are needed. By creating several learning cycles, we provided extensive possibilities for learners to develop the metacognitive skills of self-monitoring and self-regulation, and we assessed the effect of this training on a transfer task.

As a further conclusion derived from the present studies, there are two reasons why acquiring metacognitive strategies should be embedded within the context of problem solving in a domain. First, studies abound that demonstrate learners' failure to transfer newly acquired (meta)cognitive strategies to other situations, even if they have passed an exam on the topic of study (Boekaerts, 1997; Blagg, 1991; Derry & Murphy, 1986; Vosniadou, 1992). Therefore, we argue that, to improve transfer, learners should not be merely instructed about the use of the metacognitive strategies, but rather be provided with ample opportunity to practice the strategies in diverse situations. In that regard, a transfer task is needed to demonstrate a possible effect of the metacognitive strategies in novel situations. Furthermore, discussing novices' capabilities to profit from self-regulated learning, Winne (1995b) postulates that learners of all levels monitor their learning activities to a certain extent, and assess relevance and appropriateness of study tactics. As long as these monitoring activities are embedded in the regular problem-solving procedures learners use, and to the extent that these problem-solving procedures are automated, the monitoring activities are not complex, and can be performed successfully by those of low prior knowledge. However, in order to be effective, the metacognitive strategies need to be tightly woven into learners' regular problem solving strategies. This is exemplified in the present studies, as learners in Chapters 4 through 6 developed metacognitive skills while simultaneously solving domain problems (i.e., predicting the correct chess moves).

As already suggested by work on the expertise reversal effect (Kalyuga et al., 2003), merely transferring the expert learning strategy to novices will not necessarily enhance learning. Instead, certain adaptations to the instruction, that respect the essence of the expert strategy, are required to maximize learning opportunities for low knowledgeable learners. For example, the results of the studies in Chapter 5 and 6 suggest that merely requiring novice learners to estimate their understanding of chess moves does not improve self-awareness, but that a generation task that activates relevant cues about understanding prior to rating comprehension is needed in order to reveal the effect of estimating understanding on learning. A similar example of this conclusion is found in recent research by Mathan and Koedinger (2005) who, rather than providing an expert model, confronted learners with an "intelligent novice model". In contrast to an expert model, the intelligent novice model recognizes the possibility of making errors, and allows for opportunities of error correction. Based on empirical studies on the effect of the intelligent novice model, Mathan and Koedinger (2005) conclude that students who worked with the intelligent novice model performed significantly better on tests of problem solving, transfer, and retention than students who received the expert model.

Directions for future research

The studies reported in this dissertation highlight a relatively new line of research in educational psychology. Several years ago, studies reporting the effect of metacognitive strategy training on low-knowledge learners were rare, but over the last couple of years, some of these have surfaced in the literature (e.g., Butler, 1998; Azevedo, 2004). Nevertheless, a lot remains to be studied in this area. To further substantiate evidence for the deliberate practice theory, the field is in need of longitudinal studies that examine the relation between deliberate practice and performance from the beginning of practice to the point of reaching expert performance. The evident methodological issues that are associated with longitudinal research have until now prevented deliberate practice studies of this kind. However, to further test the deliberate practice assumptions in low-knowledge learners, longitudinal research is needed that focuses on individuals who recently started practice in a domain, and follows their performance development over time. Moreover, in line with Ericsson and Smith's (1991) description of the empirical study of expertise, after having determined that deliberate practice mediates expert performance, experimental research is needed that examines how deliberate practice can be performed effectively in different domains.

Furthermore, research is needed that studies the learning strategies of novices and that attempts to optimize their learning gains. Mainly when novices who have no prior knowledge whatsoever are concerned, little research exists that has examined how instruction can be designed for this particular group. However, given that teachers are often confronted with what we term "absolute novices" (i.e., in music, sports, physics, chemistry, foreign language learning), research in this group of learners is highly relevant to education.

The studies undertaken in this dissertation lead to more specific directions for future research as well. For example, concerning our research on the effect of fostering metacognitive processes in novices, the studies described in this dissertation tested transfer of acquired skills, but only with a short-term transfer task. From an educational perspective, it would be interesting to examine to what extent these skills remain useful after a more extended period of time. Future research should take into account a long-term transfer task as well. Furthermore, our studies were limited given that the effect of the procedures that intended to foster metacognitive activities was tested only by examining understanding of the chess principles as exemplified by correct move predictions in the learning phase and performance on a transfer task. This design of study was chosen as we were mainly interested in determining whether stimulating metacognitive activities during learning would foster discovery of chess principles. Therefore, we did not examine spontaneous transfer of these metacognitive activities to a similar, but novel learning environment that did not explicitly require learners to self-monitor or self-regulate. In further research, correct and spontaneous use of these metacognitive skills in novel learning situations should also be assessed. Finally, mainly for reasons of external validity, our studies were all situated within the domain of chess. To determine the generalizability of our findings to other cognitive skill domains and further extend insights on how to stimulate metacognitive processes in novice learners, there is a call for research that tests the instructions developed in this dissertation in other domains outside chess.

Chapter 8

Samenvattting en discussie

De studies in dit proefschrift zijn uitgevoerd met als doel een nieuw licht te werpen op een aantal kwesties dat momenteel aan de orde is in onderzoek naar de ontwikkeling van expertise, in het bijzonder in het onderzoek naar het bevorderen van metacognitieve processen bij lerenden. In het algemeen zijn de studies die hier beschreven worden uitgevoerd om te bepalen in welke mate novieten kunnen leren van metacognitieve processen zoals die door experts worden gebruikt. Uitgaande van leer- en instructieprincipes in de theorie van "deliberate practice" (doelbewuste oefening) van Ericsson (Ericsson et al., 1993) had ons onderzoek tot doel om de implicaties van deze theorie te testen bij novieten of niet-experts. Naar aanleiding van een analyse van de oefenactiviteiten van experts beweren Ericsson en collega's (1993) dat exceptionele prestaties niet het resultaat zijn van een specifieke genetische predispositie, maar toegeschreven kunnen worden aan jaren van intensieve oefening. Met name wanneer oefening bewust is en zich concentreert op taken die ontworpen zijn om prestatie te verbeteren (in tegenstelling tot taken die voor het plezier uitgevoerd worden), zal de ontwikkeling van expertise gemaximaliseerd worden. Specifieker geformuleerd, wanneer de leertaak van een geschikte moeilijkheidsgraad is, wanneer de leerlingen informatieve feedback krijgen over hun prestaties, en wanneer er voldoende ruimte is voor herhaling en verbetering van fouten spreken we van "deliberate practice" (doelbewuste oefening). Ericsson en collega's toonden aan dat het totale aantal uren dat individuen besteed hadden aan doelbewuste oefening direct bepalend was voor hun uiteindelijke prestatieniveau. Zij observeerden bij expert musici een monotoon stijgende relatie tussen de totale hoeveelheid "deliberate practice" en prestatieniveau. Deze monotone relatie tussen doelbewuste oefening en prestatie is sindsdien meerdere malen getoetst en bevestigd op verschillende terreinen (e.g., schaken, Charness et al., 1996, 2005; sport, Helsen et al., 1998; Hodge & Deakin, 1998; Hodges et al., 2004). De conclusie uit deze studies was dat, in plaats van aangeboren talent, doelbewuste oefening een centrale rol speelde bij het bereiken van expertniveau.

Aangezien er inmiddels breed bewijs is vergaard voor de relatie tussen "deliberate practice" en de prestaties van experts, zou een volgende stap in onderzoek zijn, om te bepalen in hoeverre deze relatie ook voor niet-experts of zelfs beginnende lerenden geldt. Oftewel: Verbeteren de prestaties van beginnende lerenden ook door "deliberate practice"? Vanuit onderwijskundig perspectief zou het interessant zijn te analyseren onder welke omstandigheden novieten kunnen profiteren van leerstrategieën die normaal gesproken als kenmerkend voor experts worden gezien. Hoewel historisch gezien theoretici doorgaans betwijfeld hebben of beginners metacognitieve vaardigheden kunnen leren (e.g., Kalyuga et al., 2003; Kanfer & Ackerman, 1989; Winne, 1995a), geven recente studies een optimistischer beeld. Deze concluderen dat onder bepaalde omstandigheden, beginnende lerenden (i.e., novieten) in staat zijn om metacognitieve strategieën succesvol toe te passen (Azevedo, 2004; Butler, 1998; Mathan & Koedinger, 2005), wanneer er rekening gehouden wordt met hun gebrek aan hiërarchisch georganiseerde schema's en hun beperkte werkgeheugencapaciteit. Daarnaast is het belangrijk dat specifieke principes uit metacognitief onderzoek toegepast worden (e.g., Koriat, 1997), waardoor een op maat gesneden instructie ontstaat die metacognitieve processen stimuleert bij novieten. Een wetenschappelijke test van een dergelijke instructie zal inzicht bieden in hoe, en op welke manier, expert leerstrategieën gestimuleerd kunnen worden bij novieten.

Overzicht van de belangrijkste bevindingen

Vooraleer het effect van het bevorderen van metacognitieve processen bij schaaknovieten te bestuderen in een experimentele setting, besloten wij eerst in een veldstudie te onderzoeken of er aanwijzingen zijn dat niet-experts in staat zijn te leren van expert strategieën als "deliberate practice". Daarom werd in de studie beschreven in **Hoofdstuk 2** voor een groep adolescente topschakers de relatie tussen "deliberate practice" en schaakprestaties geanalyseerd vanaf het begin van hun schaakoefening, gebruik makend van een "linear mixed models analysis". Deze analyse stelde ons in staat om de bijdrage van "deliberate practice" aan het prestatieniveau te bepalen, niet alleen voor het huidige niveau, maar ook voor het niveau op vroegere momenten in hun schaakcarrière. Daarnaast werd deze studie uitgevoerd om één van de belangrijkste kritiekpunten op de "deliberate practice" theorie te toetsen, welke stelt dat de factoren talent en "deliberate practice" in eerder onderzoek met elkaar verward zouden kunnen zijn (Sternberg, 1996). Wanneer alleen gekeken wordt naar individuen die uiteindelijk expertniveau bereiken, dan is het mogelijk dat de "deliberate practice" theorie gebaseerd is op een deel van de populatie dat een predispositie heeft om effectief doelbewust te oefenen. Deze aanname kan getoetst worden door te bestuderen in hoeverre investeringen in "deliberate practice" eenzelfde effect op prestatieverbetering hebben voor individuen die expertniveau bereikt hebben en voor individuen die op enig moment gestopt zijn. In Hoofdstuk 2 werd deze vergelijking gemaakt tussen een groep adolescente topschakers die nog steeds deelnamen aan de centrale training van de Koninklijke Nederlandse Schaakbond (KNSB) en een groep schakers die op enig moment vrijwillig met deze training zijn gestopt. De schaakprestaties van deze groep werden gemeten aan de hand van de schaakratings die de KNSB verstrekt.

De resultaten lieten zien dat vanaf de leeftijd waarop de schakers serieus schaakten, “deliberate practice” een betrouwbare voorspeller van hun schaakratings was. Het model dat “deliberate practice” en hun fluctuaties over tijd bevatte, verklaarde meer dan 70% van de variantie in schaakprestaties. Eerdere studies die niet de factor tijd meenamen, konden maximaal 60% van de variantie verklaren (e.g., Charness et al., 1996, 2005; Hodges et al., 2004). Deze resultaten impliceren dat zelfs wanneer individuen op lager niveau presteren (in dit geval, op eerdere momenten in hun schaakcarrière) “deliberate practice” een duidelijke voorspeller van schaakprestatie is. Bovendien, voor zowel de doorzetters als de uitvallers hadden gelijke investeringen in “deliberate practice” ook een vergelijkbaar effect op prestatie. De groepen doorzetters en uitvallers verschilden dus niet zozeer van elkaar in de mate waarin zij van “deliberate practice” profiteerden, maar in de hoeveelheid tijd die ze aan doelbewuste oefening besteedden. Deze bevindingen spreken Sternbergs (1996) assumptie tegen dat experts gekarakteriseerd kunnen worden door een talent om van “deliberate practice” te profiteren.

Als laatste bestudeerden we in Hoofdstuk 2 de mogelijke invloed van “deliberate practice” op prestatieverschillen tussen mannen en vrouwen in schaken. Onze analyse bevestigde dat mannen doorgaans hogere schaakprestaties leveren dan vrouwen, maar tevens dat dit verschil voornamelijk toe te schrijven is aan variatie in de tijd die mannen en vrouwen aan “deliberate practice” besteden. Onze resultaten lieten zien dat mannen gemiddeld gezien meer tijd aan “deliberate practice” besteden dan vrouwen.

Eerder onderzoek heeft de mogelijk essentiële rol die motivatie speelt om “deliberate practice” uit te kunnen voeren grotendeels links laten liggen. Daarom werd in de studie beschreven in **Hoofdstuk 3** onderzocht hoe algemene prestatiemotivatie en schaakspecifieke motivatie een invloed uitoefenen op de totale hoeveelheid “deliberate practice” en het huidige prestatieniveau. Hoewel experimenteel onderzoek heeft laten zien dat het effect van competitie op intrinsieke motivatie valt te betwijfelen (e.g., Deci et al., 1981; Reeve et al., 1985; Vallerand & Reid, 1984), suggereren toegepaste studies in sport dat competitiviteit kenmerkend zou kunnen zijn voor individuen die op uitzonderlijk niveau presteren (e.g., Brewer et al., 1993; Gould et al., 2002). Om deze kwestie nader te bestuderen, onderzocht de studie in Hoofdstuk 3, aan de hand van structural equation modelling, in hoeverre een competitieve prestatieoriëntatie een invloed heeft op tijd die individuen besteden aan “deliberate practice”. Onze resultaten gaven aan dat doorzetters een hogere prestatiemotivatie hadden dan uitvallers, en dat schaakspecifieke motivatie en prestatiemotivatie beide een duidelijke invloed uitoefenden op tijd die individuen aan “deliberate practice” besteden.

Een model waarin motivatie als het resultaat in plaats van de oorzaak van "deliberate practice" en prestatie werd gezien, paste minder goed op de onderliggende datastructuur. Deze studie gaf inzicht in één van de noodzakelijke voorwaarden voor bewuste oefening die aandacht verdient in training en onderwijs. Aangezien onze resultaten lieten zien dat wanneer de motivatie daalt ook de investeringen in "deliberate practice" dalen, kunnen we hieruit concluderen dat trainers en leraren regelmatig de motivatie van leerlingen zouden moeten toetsen, teneinde "deliberate practice" vol te kunnen houden.

Nadat we vastgesteld hadden dat bewuste oefening ook voor individuen met een lager kennisniveau positieve effecten op leren kan hebben, besloten we in de laatste 3 studies specifieker te onderzoeken wat het effect is van het stimuleren van metacognitieve processen op de leerprestaties van schaaknovieten. In **Hoofdstuk 4** werden schaaknovieten gevraagd verschillende computervoorbeelden van het schaakeindspel Koning en Toren tegen Koning te bestuderen onder één van drie experimentele condities. De eerste groep werd gevraagd de zetten van de computer te voorspellen, terwijl ze tegelijkertijd hun voorspellingen hardop aan zichzelf uitlegden (i.e., zelfverklaringen geven). De tweede groep voorspelde de zetten van de computer zonder zelfverklaringen te geven. De derde groep observeerde alleen de zetten van de computer zonder te voorspellen of zelfverklaringen te genereren. Onze onderzoekshypothese luidde dat de groep die zowel de zetten van de computer voorspelde als zelfverklaringen genereerde uiteindelijk beter zou presteren dan de groep die alleen voorspelde, aangezien de eerste groep expliciet gestimuleerd werden om te reflecteren op de verklaringen voor hun voorspellingen. Ook vermoedden wij dat de groep die alleen voorspelde beter zou presteren dan de groep die de zetten van de computer observeerde, omdat deze laatste groep in het geheel niet aangezet werd tot reflectie. Onze resultaten toonden dat de groep die voorspelde en zelfverklaringen gaf in de leerfase de principes van het eindspel sneller afleidde en bovendien in de testfase vaker in staat was om de Koning schaakmat te zetten dan de twee andere groepen. Er werden geen verschillen gevonden tussen de twee overige groepen. Deze bevindingen suggereren dat, ondanks de hoge werkgeheugenbelasting, het geven van zelfverklaringen tijdens het leren van een complexe vaardigheid als schaken een positief effect kan hebben op de leerprestaties. Alleen voorspellen zonder zelfverklaringen leidt waarschijnlijk niet tot dezelfde taakconcentratie die nodig is om de juistheid van de afgeleide principes te evalueren. Wanneer proefpersonen gevraagd werden te voorspellen zonder zelfverklaringen te geven, bleef hun zelfevaluatie op een oppervlakkiger niveau, waardoor hun leerprestaties vergelijkbaar waren met diegene die alleen de zetten van de computer geobserveerd hadden.

De studie in Hoofdstuk 4 liet zien dat de leerprestaties van novieten toenemen wanneer zelfreflectie gestimuleerd wordt door proefpersonen te vragen zelfverklaringen te geven. De studies gerapporteerd in Hoofdstuk 5 en 6 werden opgezet om specifieker bepaalde metacognitieve activiteiten te stimuleren en hun effect op leren te bestuderen. Gebaseerd op eerder onderzoek met betrekking tot het memoriseren van woordparen (e.g., Son & Metcalfe, 2000; Thiede, 1999; Thiede & Dunlosky, 1999) zijn self-monitoring (hierna "zelfevaluatie" genoemd) en zelfregulatie geïdentificeerd als essentiële metacognitieve vaardigheden die een gedocumenteerd effect op geheugenprestatie blijken te hebben. Hoewel studies op het gebied van het memoriseren van woordparen talrijk zijn, is er tot nu toe geen onderzoek gedaan naar het effect van zelfevaluatie en zelfregulatie bij het leren van complexe, cognitieve vaardigheden. In de studie in **Hoofdstuk 5** werd zelfevaluatie daarom geoperationaliseerd door schaaknovieten te vragen om, na het bestuderen van een voorbeeld van het eindspel van Koning en Dame tegen Koning, in te schatten hoe zeker zij waren dat ze in de toekomst soortgelijke zetten juist zouden kunnen voorspellen. Deze schatting (vaak een "Judgment of Learning" of JOL genoemd) werd vervolgens gecorreleerd met hoe goed ze de zet daadwerkelijk voorspelden (score 0, 1 of 2). Een hoge, positieve correlatie tussen deze maten was een teken van accurate zelfevaluatie. Zelfregulatie werd berekend door proefpersonen, na het geven van hun JOLs, te vragen of ze nog een voorbeeld van de zetten wilden bestuderen (ja = 1, nee = 0), en die score te correleren met de JOLs. Een hoge, negatieve correlatie was een teken van adequate zelfregulatie: proefpersonen kozen die zetten waarvan zij vonden dat ze ze slecht begrepen hadden. Hierbij werd verondersteld dat adequate zelfregulatie voortvloeide uit accurate zelfevaluatie: Alleen wanneer proefpersonen een goede inschatting van hun eigen begrip maken, zullen deze inschattingen ook een waardevolle basis voor hun herselecties zijn. Wanneer dit niet het geval is, zullen proefpersonen zetten selecteren die zij wel begrepen hadden, of bepaalde zetten juist niet selecteren, terwijl die slecht begrepen waren.

De studie uit Hoofdstuk 5 bevatte twee onafhankelijke variabelen, elk met twee niveaus, waardoor in totaal 4 groepen gevormd werden. De eerste onafhankelijke variabele bepaalde hoe het genereren van JOLs een effect had op metacognitieve processen en leerprestaties. Na het bestuderen van een eindspel van Koning en Dame tegen Koning werd aan twee groepen gevraagd voor iedere zet van wit aan te geven hoe zeker ze waren dat ze een soortgelijke zet goed zouden kunnen voorspellen (de JOL). Om te bepalen hoe accuraat hun schattingen waren, werd de kwaliteit van hun voorspellingen van de zetten gescoord en gecorreleerd met de JOLs. De twee overige groepen werden niet gevraagd om JOLs te geven. Onze hypothese luidde dat de groep die gevraagd werd om JOLs te genereren een hogere mate van begrip van de zetten zou hebben, en daarom beter zijn leren zou kunnen zelfreguleren. Dit zou uiteindelijk tot sneller en beter begrip van de schaakprincipes leiden. Daarnaast werd aan twee groepen gevraagd minimaal twee zetten te selecteren voor herstudie, terwijl twee andere groepen vrij waren in het aantal herstudies. Gegeven de beperkte ervaring die de proefpersonen hadden met schaken zou het mogelijk zijn dat ze overschatten hoe goed ze de zetten begrepen hadden en daarom geen zetten selecteerden voor herstudie. Wij veronderstelden dat de groepen die gedwongen werden tot herselectie meer tot reflecteren werden aangezet en ook betere zelfregulatie zouden vertonen.

Onze resultaten lieten zien dat het genereren van JOLs een negatief effect had op het voorspellen van de juiste zetten in de leerfase, en dat gedwongen worden tot het selecteren van minimaal twee zetten voor herstudie een positief effect had op de kwaliteit van de voorspellingen in de leerfase, zelfs wanneer er gecontroleerd werd voor het daadwerkelijke aantal herselecties. Tijdens de testfase presteerden de gedwongen selectie groepen beter dan de vrije selectie groepen door vaker de Koning schaakmat te zetten. Er werden in de testfase geen verschillen geconstateerd tussen de groepen die wel en niet JOLs genereerden. Ook bleek dat de groepen die gedwongen werden tot herselectie meer accurate JOLs maakten en betere zelfregulatie toonden dan de JOL groepen. Blijkbaar geldt voor deze proefpersonen met beperkte voorkennis dat het genereren van JOLs metacognitieve processen niet verbetert en tot op zekere hoogte het leren van de eindspel principes zelfs in de weg staat. Daarentegen heeft het dwingen tot minimale herselectie van twee zetten voor herstudie, een positief effect op zelfevaluatie en zelfregulatie en dientengevolge ook op leerprestaties. Deze bevindingen worden in Hoofdstuk 5 bediscussieerd in het licht van het “expertise reversal effect” (Kalyuga et al., 2003). De verklaring die daar gegeven wordt benadrukt dat leerstrategieën die experts gebruiken niet zondermeer geschikt zijn voor novieten en vice versa, als gevolg van verschillen in kennisorganisatie. Aanpassing van expertstrategieën naar de context van novieten is nodig om leerprocessen van individuen die geen ervaring met het onderwerp hebben te maximaliseren.

Om verder te onderzoeken hoe metacognitieve vaardigheden bevorderd kunnen worden in individuen met weinig voorkennis, beschrijft de studie in **Hoofdstuk 6** het effect van een aangepaste versie van de experimentele procedure zoals getest in Hoofdstuk 5. Alvorens deze vernieuwde procedure experimenteel te toetsen, vergeleek het eerste experiment van dit Hoofdstuk de kwaliteit van zelfevaluatie en zelfregulatie tussen schakers met en zonder eerdere ervaring met het eindspel Koning-Dame tegen Koning. Dit experiment werd uitgevoerd met als doel te bepalen in hoeverre het effectieve gebruik van metacognitieve vaardigheden afhankelijk is van eerdere ervaring met het domein. Onze resultaten gaven aan dat zelfs voor de ervaren schakers het niet vanzelfsprekend was om de metacognitieve vaardigheden correct uit te voeren. Ook meer ervaren schakers hebben dus expliciete oefening nodig om deze vaardigheden toe te passen. In het tweede experiment werd dan ook een aangepaste versie van de instructie die gebruikt was in Hoofdstuk 5 ontworpen die erop gericht was de metacognitieve vaardigheden van onervaren lerenden te stimuleren. Gebaseerd op Koriats (1997) "cue-utilization framework" werden proefpersonen gestimuleerd, direct voordat ze een JOL moesten genereren, relevante informatie over hun begripsniveau uit hun geheugen te activeren. De relevante informatie over hun begripsniveau activeerden zij door nogmaals de zet van de computer te voorspellen, maar nu zonder dat ze feedback van de computer kregen. Omdat ze direct daarna een JOL moesten genereren, was de informatie over hun begripsniveau op dat moment nog actief. Om verder de werkgeheugenbelasting te minimaliseren, werd de selectiefase direct na de JOL fase geplaatst. Proefpersonen hoefden dus niet langer de relevante informatie over begrip op te halen uit het geheugen, aangezien deze nog actief was in het werkgeheugen. Hoewel deze "generatie taak" een positief effect had op de kwaliteit van zelfevaluatie en op de prestatie op de testtaak, bleek zelfregulatie niet te verbeteren door deze instructie.

In het derde en laatste experiment van Hoofdstuk 6, besloten we verder de werkgeheugenbelasting die veroorzaakt wordt door de metacognitieve taken terug te dringen. In experiment 1 en 2 boden we proefpersonen een tamelijk complexe leeromgeving: Lerenden moesten niet alleen analyseren in hoeverre hun voorspellingen overeenkwamen met de computerzetten maar ook afleidden in welke mate hun JOLs correct waren door die te vergelijken met de kwaliteit van hun zetvoorspelling. Deze complexe taak heeft waarschijnlijk tot gevolg gehad dat het tweede deel van de taak niet of niet correct uitgevoerd werd. Om het eerste deel van de taak eenvoudiger te maken, kregen proefpersonen in het derde experiment feedback over de kwaliteit van hun voorspellingen, zodat hun informatieverwerking zich kon richten naar het evalueren van de correctheid van hun JOLs. Er werd een prestatiescore (tussen 1 en 4, 1 = slechte zet, 4 = beste zet mogelijk) toegekend aan iedere zetvoorspelling. Omdat deze feedback direct voordat proefpersonen een JOL moesten genereren over de zet werd gegeven, werden omstandigheden voor vergelijking tussen feedback en JOL geoptimaliseerd. Onze hypothese luidde dat door de feedback verwerkingscapaciteit beschikbaar werd die gebruikt zou worden om metacognitieve vaardigheden als zelfevaluatie en zelfregulatie te ontwikkelen. In tegenstelling tot onze verwachtingen, lieten onze resultaten zien dat het effect van feedback op zelfevaluatie beperkt was, en dat zelfregulatie zelfs slechter was dan in de groep die geen feedback kreeg. Een recente studie van Mathan en Koedinger

(2005) suggereert dat directe feedback de ontwikkeling van zelfevaluatieve vaardigheden zou kunnen hinderen, omdat lerenden niet langer gedwongen worden tot reflectie maar de feedback van de computer of tutor af kunnen wachten. Het is belangrijk om te melden dat de feedback wel een positief effect op de testprestatie had, wat aangeeft dat de feedback het leren van de eindspelprincipes bevorderde.

Discussie

In de literatuur is veel geschreven over de definiërende kenmerken van effectieve zelfregulatie (Butler & Winne, 1995). Naar onze mening zijn deze kenmerken voornamelijk interessant vanuit het oogpunt van de docent/trainer die deze zou willen onderwijzen. Anders gezegd, het onderzoek naar de metacognitieve strategieën die bijvoorbeeld door experts gebruikt worden, zou een voedingsbodem kunnen zijn voor onderzoek naar hoe deze het beste gestimuleerd kunnen worden in niet-zelfreguleerders. Een typisch voorbeeld van een dergelijke onderzoekscyclus vinden we bijvoorbeeld in het onderzoek naar het zelfverklaringseffect (Chi et al., 1989). Na eerst het positieve effect van spontane zelfverklaringen op leren te hebben vastgesteld, onderzochten Chi en collega's in welke mate het instrueren van leerlingen in het geven van zelfverklaringen tijdens het bestuderen van een tekst het begrip van de tekst bevorderde (Chi et al., 1994). Aangezien deze instructie inderdaad een effect had op begrip van de tekst over de bloedsomloop, had deze studie relevante implicaties voor training en onderwijs. De studies in dit proefschrift volgen een soortgelijke empirische cyclus. Als eerste werd in Hoofdstuk 2 en 3 het effect van expert leerstrategieën (i.e., "deliberate practice") bepaald. In de resterende hoofdstukken 4 tot en met 6 wordt het effect van het bevorderen van deze metacognitieve processen onderzocht bij verschillende groepen van beginnende schakers. In deze onderzoekscyclus wordt eerst een empirische basis voor het gebruik van een bepaalde techniek gevormd door het spontane gebruik ervan te analyseren. Wanneer het positieve effect van de techniek op die manier is aangetoond, worden mogelijkheden voor instructie van de techniek getoetst bij proefpersonen die de techniek niet spontaan gebruiken. De resultaten van deze studies leiden uiteindelijk tot een aantal implicaties van theoretische en onderwijskundige aard. Voor de studies die in dit proefschrift gerapporteerd zijn worden deze implicaties beneden beschreven.

Hoewel het geen expliciet doel was van dit onderzoek, lijken de huidige studies op een bepaalde manier een omgevingspositie in te nemen in het nature-nurture debat over verklaringen voor uitzonderlijke prestaties (voor een overzicht, zie Howe, Davidson, & Sloboda, 1998). De studies in Hoofdstuk 2 en 3, bijvoorbeeld, benadrukken de noodzaak van uitgebreide domein-specifieke ervaring om expert in schaken te worden. Schaak-specifieke ervaring, vooral wanneer het "deliberate practice" betrof, bleek een betrouwbare voorspeller van prestatie te zijn onder jeugdige topschakers. Factoren als geslacht en doorzetten/uitvallen, droegen slechts marginaal bij aan dit verband. Hoewel schaakspecifieke motivatie en algemene prestatiemotivatie een grote invloed bleken te hebben op "deliberate practice", zijn er geen redenen om aan te nemen dat deze vormen van motivatie aangeboren zijn en niet door omgevingsfactoren beïnvloed worden. Onze resultaten dragen bij aan het uitgebreide bewijs dat uitzonderlijke prestaties, in welke vorm dan ook, het gevolg zijn van jarenlange domein-specifieke oefening (e.g., Ericsson, Chase, & Faloon, 1980; Ericsson & Polson, 1988; Gordon, Valentine, & Wilding, 1984; Howe, Davidson, & Sloboda, 1998). Deze studies, evenals Hoofdstuk 2 en 3 van dit proefschrift, leiden impliciet tot de conclusie dat variatie in prestatie als resultaat van oefening vele malen groter is dan variatie in prestatie die ontstaat enkel en alleen als gevolg van individuele, aangeboren verschillen. Ook al bestaan er duidelijke verschillen tussen mensen in genetische aanleg, verschillen in de hoeveelheid tijd die besteed wordt aan "deliberate practice" hebben een veel groter effect op prestatie.

De studies in Hoofdstuk 4, 5 en 6 vullen de aanname van domein-specificiteit op twee manieren verder aan. Als eerste laten deze studies zien dat de relatie tussen domein-specifieke ervaring en prestatie ook geldt voor individuen van lagere kennisniveaus. Na intensieve oefening die specifiek ontworpen was om expertiseontwikkeling te bevorderen, bleken absolute novieten in het schaken in staat om de principes van een schaakeindspel af te leiden en correct toe te passen. Daarnaast geven onze studies aan dat het *domeinspecificiteitsprincipe* ook van toepassing is op de ontwikkeling van metacognitieve vaardigheden als zelfevaluatie en zelfregulatie. De domein-specifieke ervaring stelde onze proefpersonen in staat deze metacognitieve processen te ontwikkelen en te gebruiken bij het oplossen van nieuwe problemen. Herhaling van deze vaardigheden tijdens de leerfase bleek de kwaliteit van de metacognitieve processen bij deze groep onervaren schakers ten goede te komen. Samenvattend concluderen wij dat de zelfregulatie vaardigheden (e.g., Butler & Winne, 1995) niet beschreven hoeven te worden als zijnde aangeboren en ongevoelig voor training. Onze studies benadrukken dat, met de juiste training, deze vaardigheden geleerd kunnen worden aan diegene die er geen gebruik van maken. Echter, aangezien het spontane gebruik van deze vaardigheden zeldzaam is en transfer naar nieuwe situaties doorgaans beperkt is, is intensieve domein-specifieke oefening noodzakelijk om de vaardigheden correct te ontwikkelen. Wij zullen een aantal conclusies bespreken dat voortvloeit uit ons onderzoek, en dat relevant is bij het ontwikkelen van training in metacognitieve vaardigheden.

Conclusies voor instructie van metacognitieve vaardigheden bij novieten

Wanneer het om training van absolute beginners gaat, is het als eerste van groot belang dat de instructie zo veel mogelijk onafhankelijk is van directe sturing door een leraar. Bij voorkeur zou de leerling, na een korte introductie, in staat moeten zijn om de metacognitieve vaardigheden zelf te oefenen zonder begeleiding van een leraar of trainer. Om uiteindelijk de kans op transfer van de geleerde strategie naar andere situaties te verhogen, is het noodzakelijk dat leerlingen hun eigen leeractiviteiten sturen. Daarnaast zal de docent-onafhankelijkheid de mogelijkheden voor onderwijskundige implementatie vergroten, aangezien het weinig investering van de docent vraagt. Ook dwingt docent-onafhankelijkheid een bepaalde mate van simpliciteit in instructie af, waardoor de instructie wellicht makkelijk gebruikt kan worden in andere domeinen. De instructie zoals ontwikkeld in schaken, zal bijvoorbeeld na kleine aanpassingen ook van toepassing zijn op het trainen van andere cognitieve vaardigheden. In de huidige studies was de docent-onafhankelijkheid met name zichtbaar in Hoofdstuk 4, waar proefpersonen zelfverklaringen gaven terwijl ze de zetten van een schaakcomputer voorspelden. Om leerlingen effectief te laten zelfverklaren, is een uitleg van een docent nodig, die bij voorkeur een goed voorbeeld hiervan geeft, maar al met al hoeft deze instructie niet langer dan 10 minuten te duren. Hoewel de instructies getest in Hoofdstuk 5 en 6 docent-onafhankelijk waren, kunnen deze instructies niet vanzelfsprekend tijdens zelfstudie gebruikt worden vanwege afhankelijkheid van een computer. De proefpersonen maakten in dit geval gebruik van een computerapplicatie waarin alle relevante informatie en oefeningen gestructureerd waren. Echter, met een aantal eenvoudige aanpassingen en beperkte instructie van een docent kunnen leerlingen de cruciale aspecten van de instructie oefenen in een zelfstudie situatie zonder gebruikmaking van een computer. Bijvoorbeeld, wanneer leerlingen geleerd wordt om de te bestuderen stof op te delen in betekenisvolle eenheden (e.g., de alinea's van een tekst, of aparte sommen in wiskunde) en zij in moeten schatten hoe goed ze elke eenheid begrepen hebben, kan hen ook geleerd worden om op basis van die schattingen te bepalen welk deel van de stof opnieuw bestudeerd zou moeten worden. Ook al hebben wij in onze studies de instructie volledig via de computer aangeboden, met enkele aanpassingen zou die vertaald kunnen worden naar een zelfstudie omgeving.

Omdat novieten gekenmerkt worden door een gebrek aan geautomatiseerde procedures voor het oplossen van problemen, is het noodzakelijk dat training gericht op het stimuleren van metacognitieve processen een zo laag mogelijk beroep doet op hun werkgeheugencapaciteit. Onnodige extra belasting van het geheugen zal ongetwijfeld leren hinderen (Sweller, 1999; Sweller, Van Merriënboer, & Paas, 1998). Dit effect observeerden wij in Hoofdstuk 5, waar beginnende schakers minder leerden wanneer hun gevraagd werd om te reflecteren over hun begrip van het schaken door JOLs te genereren. Pas toen we specifieke voorzorgsmaatregelen namen om de werkgeheugenbelasting te verminderen (Hoofdstuk 6), werd een positief effect van deze metacognitieve processen op leren zichtbaar.

Gegeven dat het genereren van zelfverklaringen ook een belasting van het werkgeheugen met zich meebrengt, lijkt het positieve effect van zelfverklaren bij beginners in contrast met het bovenstaande. Echter, in tegenstelling tot de metacognitieve activiteiten die in Hoofdstuk 5 getest werden, worden zelfverklaringen volledig gestuurd door de proefpersonen. Zij werden geïnstrueerd om continu zelf te verklaren, maar de inhoud en de uitgebreidheid van de zelfverklaringen werden uiteindelijk bepaald door proefpersonen, afhankelijk van de beschikbare verwerkingscapaciteit. Wanneer een proefpersoon moeite had met een bepaalde zet, was er waarschijnlijk maar beperkte ruimte voor zelfverklaringen beschikbaar, waardoor deze in aantal afnamen. Het positieve effect van zelfverklaringen op leren boven verklaringen van een leraar ontstaat waarschijnlijk doordat ze volledig gestuurd worden door de leerlingen en per definitie aangepast zijn aan het niveau van het individu. Dit verkleint de kans op werkgeheugenoverbelasting (voor een overzicht van de voor- en nadelen van zelfverklaringen en verklaringen door een leraar, zie Renkl, 2002).

Ook vanuit het oogpunt van de beperkte werkgeheugencapaciteit is het cruciaal dat de metacognitieve vaardigheden door herhaalde oefening geautomatiseerd kunnen worden. Zoals al geconstateerd door Boekaerts (1997) wordt een groot deel van de metacognitieve strategieën die experts gebruiken niet bewust en gecontroleerd gebruikt, maar automatisch en zonder aandacht. Alleen daarom kunnen deze strategieën effectief zijn in veeleisende en complexe leeromgevingen. Ook bij het leren van metacognitieve vaardigheden aan beginners, is het dus van belang dat er voldoende ruimte is voor automatisering. In de huidige studies werden er steeds meerdere leercycli aangeboden, waardoor proefpersonen meerdere malen in staat werden gesteld de metacognitieve vaardigheden te oefenen.

Een verdere conclusie uit de huidige studies is dat er twee redenen zijn waarom het leren van metacognitieve vaardigheden ingebed moet zijn in de context van het oplossen van problemen. Als eerste zijn er diverse studies die aantonen dat in het algemeen transfer van pasgeleerde (meta)cognitieve vaardigheden naar een andere situatie laag is (Boekaerts, 1997; Blagg, 1991; Derry & Murphy, 1986; Vosniadou, 1992). Om deze transfer plaats te laten vinden, stellen wij dat proefpersonen niet alleen geleerd moet worden hoe de metacognitieve vaardigheden te gebruiken, maar ook dat ze uitgebreide oefening moeten krijgen in het toepassen van deze vaardigheden in diverse situaties. Hieruit vloeit voort dat een transfer taak aan het einde van de oefensessie nodig is, om te kunnen bepalen in hoeverre proefpersonen in staat zijn de vaardigheden in andere domeinen effectief te gebruiken. Als tweede stelt Winne (1995b) dat leerlingen van alle expertiseniveaus tot op bepaalde hoogte in staat zijn om te reflecteren over hun studieactiviteiten. Wanneer deze reflectieve activiteiten ingebed zijn in de reguliere procedures die leerlingen gebruiken om problemen op te lossen, en zolang deze procedures geautomatiseerd zijn, hoeven de reflectieve activiteiten niet complex te zijn, en kunnen ze zelfs door individuen met geringe voorkennis uitgevoerd worden. Voorwaarde is dat de reflectieve activiteiten nauw verweven zijn met de probleemoplosprocedures. Ook in de huidige studies wordt dit ondersteund, aangezien proefpersonen terwijl ze de metacognitieve vaardigheden oefenden ook problemen oplosten (e.g., door het voorspellen van de zetten van de computer).

Het “expertise reversal effect” (Kalyuga et al., 2003) beschreef reeds hoe expert leerstrategieën niet direct gebruikt kunnen worden door beginnende lerenden. Bepaalde aanpassingen, die de essentie van de strategieën in tact laten, zijn noodzakelijk om beginners te kunnen laten profiteren van expert strategieën. De resultaten van Hoofdstuk 5 en 6 bijvoorbeeld, laten zien dat enkel en alleen vragen aan proefpersonen om hun begrip van de schaakzetten te schatten geen effect heeft op zelfevaluatie en leerprestatie, maar dat hiervoor een genereertaak nodig is die proefpersonen dwingt relevante informatie over begripsniveau te activeren. Een vergelijkbaar voorbeeld is recent gegeven door Mathan en Koedinger (2005) die geen expert model toonden aan hun proefpersonen, maar een “intelligent novice model”. In tegenstelling tot het expert model erkent het noviet model de mogelijkheid van het maken van fouten, en biedt daarom mogelijkheden tot verbetering hiervan. Mathan en Koedinger concluderen dat proefpersonen die met het noviet-model gewerkt hadden, beter waren in het oplossen van de problemen die ze bestudeerd hadden.

Aanbevelingen voor vervolgonderzoek

De studies die besproken worden in dit proefschrift zijn uitgevoerd vanuit een relatief nieuwe tak van onderzoek in de onderwijspsychologie. Een aantal jaren geleden waren studies naar het vermogen van beginnende lerenden om te kunnen profiteren van metacognitieve vaardigheden zeldzaam, maar de laatste jaren heeft een aantal van deze studies hun weg naar de literatuur gevonden (e.g., Butler, 1998; Azevedo, 2004). Desalniettemin is er nog veel onderzoek nodig op dit terrein om tot een duidelijk beeld van de mogelijkheden van de beginnende lerende te komen. Om verder bewijs te kunnen leveren voor de "deliberate practice" theorie, is er behoefte aan longitudinaal onderzoek dat de relatie tussen "deliberate practice" en prestatie vanaf het begin van oefening tot het bereiken van expertniveau in kaart brengt. De evidente methodologische bezwaren die aan dit type studies kleven -het duurt immers tien jaar alvorens iemand expert is-, hebben er tot op heden voor gezorgd dat dit type studies niet uitgevoerd is. Echter, ook om de invloed van "deliberate practice" op beginnende lerenden te kunnen bepalen is longitudinaal onderzoek op dit terrein essentieel. Nu het effect van "deliberate practice" op prestatie in experts is aangetoond, is het, zoals Ericsson en Smith (1991) al opmerkten, zaak om deze relatie middels experimentele studies in niet-experts te onderzoeken.

Daarnaast is onderzoek gewenst dat de leerstrategieën van novieten verder bestudeert met als doel hun leerprestaties te optimaliseren. Onderzoek naar novieten die geen enkele vorm van voorkennis bezitten, blijft schaars. Echter, aangezien docenten regelmatig geconfronteerd worden met leerlingen die volledig nieuw in een domein zijn (denk aan muziek, sport, scheikunde, natuurkunde, vreemde talen), zou onderzoek naar deze groep relevante implicaties voor onderwijs op kunnen leveren.

De studies in dit proefschrift leiden tenslotte nog tot een aantal specifieke aanbevelingen voor vervolgonderzoek. Wat betreft het onderzoek naar het stimuleren van metacognitieve vaardigheden in beginners, hebben wij in onze experimenten alleen gebruik gemaakt van een korte termijn transfer taak. Vanuit onderwijskundig oogpunt zou het interessant zijn te bestuderen in hoeverre de geleerde vaardigheden ook op de langere termijn behouden blijven. In vervolgonderzoek zou daarom een lange termijn transfer taak meegenomen moeten worden. Daarnaast bepaalden onze studies het effect van de metacognitieve vaardigheden door te analyseren hoe goed proefpersonen de schaakprincipes gebruikten tijdens het voorspellen van de zetten en het spelen van het eindspel. Deze opzet was gekozen omdat wij voornamelijk geïnteresseerd waren in welke mate het stimuleren van metacognitieve processen een effect heeft op het ontwikkelen van begrip van de domeinprincipes. Mede daarom hebben wij niet gekeken naar het spontane gebruik van de metacognitieve vaardigheden in nieuwe leersituaties. In vervolgonderzoek zou spontaan, correct gebruik van de vaardigheden in nieuwe taken bestudeerd moeten worden om hier uitsluitsel over te geven. Als laatste zijn onze studies, om redenen van externe validiteit, uitgevoerd in het schaakdomein. Om de generaliseerbaarheid van onze bevindingen vast te stellen en verder inzicht te geven in de metacognitieve vaardigheden waartoe beginners in staat zijn, is er behoefte aan onderzoek dat de instructies die hier ontwikkeld zijn in andere domeinen toetst.

References

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition*, *9*, 422–433.

Ainsworth, S., & Loizou, A. Th. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27,* 669-681.

Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: the interplay of cognitive, motivational, and strategic forces. *Advances in Motivation and Achievement, 10,* 213-250.

Alexander, P. A., Murphy, P. K., & Woods, B. S. (1996). Of squalls and fathoms: Navigating the sea of educational innovation. *Educational Researcher, 25*(3), 31-36, 39.

Alexander, P. A., Pate, P. E., Kulikowich, J. M., Farrell, D. M., & Wright, N. L. (1989). Domain-specific and strategic knowledge: Effects of training on students of differing ages or competence levels. *Learning and Individual Differences, 1*(3), 283-325.

Ames, C. (1992). Achievement goals, motivational climate, and motivational processes. In G. Roberts (Ed.), *Motivation in Sport and Exercise* (pp. 161-176). Champaign, IL: Human Kinetics Publishers.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4,* 167-207.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide.* Chicago, IL: SmallWaters Corporation.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*(4), 199-209.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523-535.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, J. L. Kamil, & P. Rosenthal (Eds.), *Handbook of Reading Research.* New York: Longman Press.

Baker, J., Côté, J., & Abernethy, B. (2003). Sport specific training, deliberate practice, and the development of expertise in team ball sports. *Journal of Applied Sport Psychology, 15,* 12-25.

Baker, J., Côté, J., & Deakin, J. (2005). Expertise in ultra-endurance triathletes early sport involvement, training structure, and the theory of deliberate practice. *Journal of Applied Sport Psychology, 17*, 64-78.

Baker, J., Deakin, J., & Côté, J. (2005). On the utility of deliberate practice: Predicting performance in ultra-endurance triathletes from training indices. *International Journal of Sport Psychology, 36*, 225-240.

Blagg, N. (1991). *Can we teach intelligence? A comprehensive evaluation of Feuerstein's Instrumental Enrichment Program.* London: Erlbaum.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: their status 20 years later. *Psychological Science, 11,* 474-480.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161-186.

Borkowski, J. G., & Peck, V. A. (1986). Causes and consequences of metamemory in gifted children. In R. J. Sternberg, & J. Davidson (Eds.), *Conceptions of giftedness* (pp.182-200). Boston: Cambridge University Press.

Brand-Gruwel, S. (1995). *Onderwijs in tekstbegrip. Een onderzoek naar het effect van strategisch lees- en luisteronderwijs bij zwakke lezers [Teaching reading comprehension].* Nijmegen, NL: Tandem Felix.

Brewer, B. W., Van Raalte, J. L., & Linder, D. E. (1993). Athletic identity: Hercules muscle or Achilles heel? *International Journal of Sport Psychology, 24*(2), 237-254.

Brown, A. (1988). Motivation to learn and understand: On taking charge of one's own learning. *Cognition and Instruction, 5,* 311-321.

Brown, J. S., & Burton, R. R. (1980). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*(2), 155-192.

Bruinsma, M. (2003). Leidt hogere motivatie tot betere prestaties? Motivatie, informatieverwerking en studievoortgang in het hoger onderwijs. [Does higher motivation result in higher achievement? Motivation, cognitive processing and achievement in higher education.] *Pedagogische Studiën, 80*(3), 226-238.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical-theoretic approach* (2nd Ed.). New York: Springer.

Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning. *Journal of Educational Psychology, 90,* 682-697.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.* MahWah, NJ: Lawrence Erlbaum Associates.

Calin-Jageman, R. J., & Horn Ratner, H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction, 23*(4), 523-543.

Carroll, W. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86,* 360-367.

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory & Cognition, 22*(5), 606-615.

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In K.A. Ericsson (Ed.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games* (pp. 51-80). Mahwah, NJ: Lawrence Erlbaum Associates.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*(2), 151-165.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Chi, M. T. H., VanLehn K., (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1,* 69–105.

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*(2), 304-316.

Côté, J., Baker, J., & Abernethy, B. (2001). Stages of sport participation of expert decision-makers in team ball sports. In A. Papaionnou, M. Goudas, & Y. Theodorakis (Eds.), *In the dawn of the new millennium – Program and proceedings of the 10th world congress of sport psychology: Vol 3* (pp. 150-152). Thessaloniki, Hellas: Christodoulidi Publishing.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition, 22*, 249-257.

Deakin, J. M., & Allard, F. (1991). Skilled memory in expert figure skaters. *Memory & Cognition*, *19*, 79–86.

De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2005a). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology, 19*, 167-181.

De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (in press). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology.*

De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2006). *The influence of achievement motivation and chess-specific motivation on deliberate practice.* Manuscript submitted for publication.

De Bruin, A.B.H., Schmidt, H.G., & Rikers, R.M.J.P. (2005b). The role of basic science and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Academic Medicine, 80*(8), 765-773.

De Bruin, A. B. H., Smits, N., Rikers, R. M. J. P., & Schmidt, H. G. (2006). *Deliberate practice predicts performance throughout time: A linear mixed models analysis.* Manuscript submitted for publication.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105-115.

Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Journal of Personality and Social Psychology, 7,* 79-83.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-determination in Human Behavior.* New York: Plenum.

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague, the Netherlands: Mouton. (Original work published 1946).

De Leeuw, N. A. (2000). *How students construct knowledge from multiple sources.* Unpublished manuscript, Learning Research and Development Center, University of Pittsburgh.

De Leeuw, N. A., & Chi, M. T. H. (2003). Self-explanation: Enriching a situation model or repairing a domain model? In G. Sinatra & P. Pintrich (Eds.), *Intentional conceptual change* (pp. 55-78). Mahwah, NJ: Lawrence Erlbaum Associates.

Derry, S. J., & Murphy, D. A. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research, 56,* 1–39.

Die, A. H., Seelbach, W. C., & Sherman, G. D. (1987). Achievement motivation, achieving styles, and morale in the elderly. *Psychology and Aging, 2*(4), 407-408.

Doll, J., & Mayr, U. (1987). Intelligenz und Schachleistung – eine Untersuchung an Schachexperten {Intelligence and performance in chess – A study of chess experts}. *Psychologische Beiträge, 29,* 270-289.

Duda, J. L. (2001). Achievement goal research in sport: pushing the boundaries and clarifying some misunderstandings. In G. Roberts (Ed.), *Advances in Motivation in Sport and Exercise* (pp. 129-182). Champaign, IL: Human Kinetics Publishers.

Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sport Psychology, 35,* 232-245.
Duncker, K. (1945). On problem solving. *Psychological Monographs, 58,* 1-113 (Whole No. 270).
Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journal of Gerontology: Psychological Sciences and Social Sciences, 52,* 178-186.
Dunn, T. G., & Shriner, C. (1999). Deliberate practice in teaching: what teachers do for self-improvement. *Teaching and Teacher Education, 15*, 631-651.
Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games* (pp. 1-50). Mahwah, NJ: Lawrence Erlbaum Associates.
Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.
Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.
Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.
Ericsson, K. A., & Polson, P. G. (1988). An empirical analysis of the mechanisms of a memory skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 305-316.
Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 1-38). New York: Cambridge University Press.
Eteläpelto, A. (1993). Metacognition and the expertise of computer program comprehension. *Scandinavian Journal of Educational Research, 37,* 243-254.
Flavell, J. H. (1987). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-236). Hillsdale, NJ: Erlbaum.
Franken, R. E., Hill, R., & Kierstead, J. (1994). Sport interest as predicted by the personality measures of competitiveness, instrumentality, expressivity, and sensation seeking. *Personality and Individual Differences, 17*(4), 467-476.
Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology, 83,* 233-235.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S., Avitissian, I., & Prosk, A. (1984). Enhancement of eyewitness memory: an empirical evaluation of the cognitive interview. *Journal of Police Science and Administration, 12*, 74-80.

Gelman, R., & Brenneman, K. (1994). First principles can support both universal and culture-specific learning about number and music. In S. A. Gelman & L.A. Hirschfeld (Eds.), *Mapping the mind: Domain specificity in language and culture* (pp. 369-390). New York, NJ: Cambridge University Press.

Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 125-186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gill, D. L. (1988). The relationship of competitiveness and achievement orientation to participation in sport and nonsport activities. *Journal of Sport & Exercise Psychology, 10,* 139-150.

Gobet, F. (1993). A computer model of chess memory. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 463–468). Hillsdale, NJ: Erlbaum.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40.

Gordon, P., Valentine, E., & Wilding, J. (1984). One man's memory: A study of a mnemonist. *British Journal of Psychology*, *75*, 1–14.

Gould, D., Dieffenbach, K., & Moffett, A. (2002). Psychological characteristics and their development in Olympic champions. *Journal of Applied Sport Psychology, 14*(3), 172-204.

Greeno, J.G. (1992). Mathematical and scientific thinking in classrooms and other situations. In D. F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 39-61). Hillsdale, NJ: Lawrence Erlbaum Associates.

Guillet, E., Sarrazin, P., Carpenter, P. J., Trouilloud, D., & Cury, F. (2002). Predicting persistence or withdrawal in female handballers with social exchange theory. *International Journal of Sport Psychology, 37*(2), 92-104.

Gustin, W. C. (1985). The development of exceptional research mathematicians. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 270-331). New York: Ballantine Books.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92,* 331-341.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in Educational Theory and Practice.* Mahwah, NJ: Erlbaum.

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education, 8,* 17-26.

Helmreich, R. L., & Spence, J. T. (1978). The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. *JSAS Catalog of Selected Documents in Psychology, 8,* 35.

Helsen, W. F., & Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport & Exercise Psychology, 20*, 12-34.

Hodge, T., & Deakin, J. M. (1998). Deliberate practice and expertise in the martial arts: The role of context in motor recall. *Journal of Sport & Exercise Psychology, 20,* 260-279.

Hodges, N. J., Kerr, T., Starkes, J. L., Weir, P. L., & Nananidou, A. (2004). Predicting performance times from deliberate practice hours for triathletes and swimmers: What, when, and where is practice important? *Journal of Experimental Psychology: Applied, 10*(4), 219-237.

Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature of expertise: A sport-specific test of Ericsson, Krampe, and Tesch-Römer's (1993) theory of 'deliberate practice'. *International Journal of Sport Psychology, 27,* 400-424.

Howard, R. (2004). Are gender differences in high achievement disappearing? A test in one intellectual domain. *Journal of Biosocial Studies, 37*, 371-380.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: reality or myth? *Behavioral and Brain Sciences, 21*(3), 399-442.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1-17.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude treatment approach to skill acquisition [Monograph]. *Journal of Applied Psychology, 74,* 657-690.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95,* 163-182.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1-24.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science, 15*(10), 661-667.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349-370.

Krampe, R. T., & Ericsson, K.A. (1996). Maintaining excellence: Deliberate practice and elite performance in younger and older pianists. *Journal of Experimental Psychology: General, 125*(4), 331-359.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963-974.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Larkin, J. H., & Reif, F., (1976). Analysis and teaching of a general skill for studying scientific text. *Journal of Educational Psychology, 72*, 348–350.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 464-470.

Long, J. D., & Long, E. W. (1987). Enhancing student achievement through metacomprehension training. *Journal of Developmental Education, 11*, 2-5.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663-679.

Maki, R. H., Serra, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84*(2), 200-210.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40*(4), 257-265.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*(1), 14-19.

McAuley, E., & Tammen, V. V. (1989). The effects of subjective and objective competitive outcomes on intrinsic motivation. *Journal of Sport & Exercise Psychology, 11,* 84-93.

McKeithen, K. B., Reitman, J., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*, 307–325.

Memon, A., & Bull, R. (1991). The cognitive interview: its origins, empirical support, evaluation, and practical implications. *Journal of Community and Applied Social Psychology, 1,* 291-307.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about Knowing.* Cambridge, MA: MIT Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Monsaas, J. A. (1985). Learning to be a world class tennis player. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 211-269). New York: Ballantine Books.

Moulaert, V., Verwijnen, M. G., Rikers, R., & Scherpbier, A. J. (2004). The effects of deliberate practice in undergraduate medical education. *Medical Education, 38*(10), 1044-1052.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: the effect of example format and generating self-explanations. *Cognition and Instruction, 16*(2), 173-199.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207-213.

Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology, 68,* 15-24.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp.1-55). Hillsdale, NJ: Erlbaum.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1166–1174.

Omar, R. Z., Wright, E. M., Turner, R. M., & Thompson, S. G. (1999). Analysing repeated measurements data: A practical comparison of methods. *Statistics in Medicine, 18*, 1587-1603.

Paas, F., & Van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86,* 122-133.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Pelletier, L. G., Fortier, M. S., Vallerand, R. J., Brière, N. M. (2001). Associations among perceived autonomy support, forms of self-regulation, and persistence: A prospective study. *Motivation and Emotion, 25*(4), 279-306.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.

Pritchard, R. D., Campbell, K. M., & Campbell, D. J. (1977). Effects of extrinsic financial rewards on intrinsic motivation. *Journal of Applied Psychology, 62,* 9-15.

Reeve, J., & Deci, E. L. (1996). Elements of the competitive situation that affect intrinsic motivation. *Personality and Social Psychology Bulletin, 22*(1), 24-33.

Reeve, J., Olson, B. C., & Cole, S. G. (1985). Motivation and performance: Two consequences of winning and losing a competition. *Motivation and Emotion, 9,* 291-298.

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*(1), 1-29.

Renkl, A. (1999). Learning mathematics from worked-out examples: analyzing and fostering self-explanations. *European Journal of Psychology of Education, 14*(4), 477-488.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12*, 529-556.

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked-out solution steps works: A cognitive load perspective. *Instructional Science, 32*(1-2), 59-82.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23,* 90-108.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2000). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52,* 453-470.

Sarrazin, P., Vallerand, R., Guillet, G., Pelletier, L., Cury, F. (2002). Motivation and dropout in female handballers: a 21-month prospective study. *European Journal of Social Psychology, 32,* 395-418.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. C. Wittrock, (Ed.), *Handbook of research on teaching* (3rd Ed.) (pp. 778-803.). New York: Macmillan.

Schönfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schönfeld, A. H. (1987). What's all the fuss about metacognition? In A.H. Schönfeld (Ed.), *Cognitive Science and Mathematics Education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611-621.

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology, 19,* 143-154.

Schroth, M. L., & Lund, E. (1994). Relationships between need achievement, sensation seeking and cognitive performance. *Personality and Individual Differences, 16*(6), 861-867.

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance.* Hillsdale, NJ: Erlbaum.

Shaft, T. M. (1995). Helping programmers understand computer programs: The use of metacognition. *Data Base Advances, 26,* 25-46.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4,* 131-135.

Shiffrin, R. M. (1996). Laboratory experimentation on the genesis of expertise. In K.A. Ericsson (Ed.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games* (pp. 337-346). Mahwah, NJ: Lawrence Erlbaum Associates.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*, 394-403.

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of deliberate practice in the development of performing musicians. *British Journal of Psychology, 87*, 287-309.

Son, L. K., & Metcalfe, J. (2000) Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221.

Spence, J. T., & Helmreich, R. L. (1983). Achievement-related motives and behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 10-74). San Francisco, CA: W.H. Freeman and Company.

Stark, R. (1998). *Lernen mit Lösungsbeispielen. Der Einfluß unvollständiger Lösungsschritte auf Beispielelaboration, Motivation und Lernererfolg* [Learning by worked-out examples. The impact of incomplete solution steps on example elaboration, motivation, and learning outcomes]. Unpublished dissertation. University of Munich, Germany.

Starkes, J. L. (1987). Skill in field hockey: The nature of the cognitive advantage. *Journal of Sport Psychology*, *9*, 146–160.

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: what is it anyway? In K. A. Ericsson (Ed.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games* (pp. 81-106). Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R. J. (1996). Costs of expertise. In K. A. Ericsson (Ed.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games* (pp. 347-354). Mahwah, NJ: Lawrence Erlbaum Associates.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4^{th} Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sweller, J. (1999). *Instructional design in technical areas.* Melbourne, Australia: ACER Press.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6*, 662-667.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129-160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66-73.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024-1037.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 31*(6), 1267-1280.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

Vallerand, R. J., & Bissonnette, R. (1992). Intrinsic, extrinsic, and amotivational styles as predictors of behaviour: a prospective study. *Journal of Personality, 60*(3), 599-620.

Vallerand, R. J., Deci, E. L., & Ryan, R. M. (1987). Intrinsic motivation in sport. In K. Pandolf (Ed.), *Exercise and Sport Science Review* (pp. 389-425). New York: Macmillan.

Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology, 6*, 94-102.

Van Gelder, T., & Bissett, M. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology, 58*(2), 142-152.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

VanLehn, K., & Jones, R. (1993). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), *Proceedings of the fifteenth annual conference of the cognitive science society* (pp. 1034-1039). Hillsdale, NJ: Erlbaum.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Yperen, N. W., & Duda, J. L. (1999). Goal orientations, beliefs about success, and performance improvement among young elite Dutch soccer players. *Scandinavian Journal of Medicine & Science in Sports, 9*(6), 358-364.

Vicente, K. J. (1988). Adapting the memory recall paradigm to evaluate interfaces. *Acta Psychologica*, *69*, 249–278.

Vicente, K. J. & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*(1), 33-57.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Journal, 41,* 347–357.

Vu, K. L., Hanley, G. L., Strybel, T. Z., & Proctor, R. W. (2000). Metacognitive processes in human-computer interaction: Self-assessments of knowledge as predictors of computer expertise. *International Journal of Human-Computer Interaction, 12*, 43-71.

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology, 93,* 557-565.

Weinert, F. E., Schrader, F. W., & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Psychology, 13*(8), 895–912.

Winne, P. H. (1995a). Inherent details in self-regulated learning. *Educational Psychologist, 30*(4), 173-187.

Winne, P. H. (1995b). Self-regulation is ubiquitous, but its forms may vary with knowledge. *Educational Psychologist, 30*(4), 223-228.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.) (pp. 153-189). Mahwah, NJ: Erlbaum.

Zimmerman, B., & Schunk, D. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Zur, O., & Gelman, R. (2004).Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly, 19*(1), 121-137.

Curriculum Vitae and Publications

Anique Bea Hubertina de Bruin was born in Maastricht on June 17th, 1977. She completed secondary education in 1995 at the Trichter College in Maastricht. In May 2000 she received a Master's degree in cognitive psychology at Maastricht University (cum laude). She started her work as a Ph. D. student at Maastricht University in April 2000. In 2001, she obtained a position as an assistant professor in psychology at the Erasmus University Rotterdam, where she became involved amongst others in the coordination of the progress test and the first year course on developmental psychology. Concurrently, she worked on research on the development of expertise in novice learners, which resulted in the present dissertation.

De Bruin, A. B. H., Rikers, R. M. J. P, & Schmidt, H. G. (2005a). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology,* 19, 167-181.

De Bruin, A. B. H., Schmidt, H. G., & Rikers, R. M. J. P. (2005b). The role of basic science and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Academic Medicine, 80* (8), 765-773.

De Bruin, A. B. H., Rikers, R. M. J. P, & Schmidt, H. G. (in press). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology.*

De Bruin, A. B. H., Van de Wiel, M. W. J., Rikers, R. M. J. P., & Schmidt, H. G. (2005c). Examining the stability of experts' clinical case processing: An experimental manipulation. *Instructional Science*, *33*, 251-270.

Rikers, R. M. J. P., & De Bruin, A. B. H. (2006). Introduction to the special issue on innovations in problem-based learning. *Advances in Health Sciences Education*, *11*(4), 315-319.