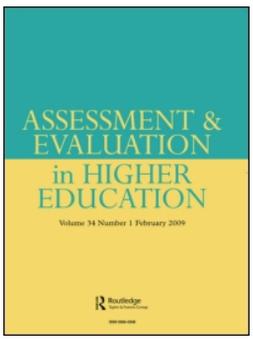

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: http://www.tandfonline.com/loi/caeh20

# Revisiting differential grading standards anno 2014: an exploration in Dutch higher education

Brian P. Godor

**To cite this article:** Brian P. Godor (2017) Revisiting differential grading standards anno 2014: an exploration in Dutch higher education, Assessment & Evaluation in Higher Education, 42:4, 596-606, DOI: 10.1080/02602938.2016.1173186

**To link to this article:** http://dx.doi.org/10.1080/02602938.2016.1173186

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 19 Apr 2016.

Submit your article to this journal

Article views: 181

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=caeh20

**Download by:** [Erasmus University] **Date:** 10 March 2017, At: 02:31

*Assessment & Evaluation in Higher Education*, 2017
Vol. 42, No. 4, 596–606, http://dx.doi.org/10.1080/02602938.2016.1173186

Routledge
Taylor & Francis Group

# Revisiting differential grading standards anno 2014: an exploration in Dutch higher education

Brian P. Godor*

*Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands*

The role that teachers have in assessing student coursework is crucial. Their 'determination' that a particular piece of student's work is 'acceptable' has many serious consequences. With a lack of debate surrounding assessment, practices may become mired in conventions and disconnected from issues such as knowledge, power and social organisation. They may also become divergent between faculties, with majors with academically stronger students grading more stringently and those with academically weaker students grading less stringently. In order to test for potential differential grading standards due to adaption-level, this study examines the relationship of pre-entry secondary school grades (English, Dutch and Mathematics) to first-year average grades in 11 faculties ($n = 3080$). The results presented demonstrate the presence of differential grading standards among the different faculties within this population.

**Keywords:** assessment; grading; differential standards; higher education; adaptation-level theory

## Introduction

> It is perhaps unfortunate, but true, that for many students grades are *the* important thing which one gets from school. (Aiken 1963, 319)

> Assessment makes more difference to the way that students spend their time, focus their effort, and perform, than any other aspect of the courses they study, including the teaching. (Gibbs 2010, 3)

The role that teachers have in assessing student coursework is crucial. Their 'determination' that a particular piece of student's work is 'acceptable' has many serious consequences. For example, the attainment of a non-sufficient grade could mean that students need to re-take the course, or even perform additional coursework to repair this deficiency. Despite this central role that grades have in students' academic careers, Boud (1990) asserts 'there has been little general questioning of how and why we assess' (101). Moreover, Hills (1991) alleges that teacher's apathy towards 'the competent use of the technical skills involved in adequate testing, evaluation, and grading' is analogous to students' 'apathy toward learning' (544). According to Delandshere (2001) there is an:

> absence of clearly articulated theories of assessment and debates regarding the philosophical stances that undergird assessment practices. While learning theories,

*Email: godor@fsw.eur.nl

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

> measurement theories and curriculum theories continue to be developed, assessment practices seem to have avoided any such formalisation even though assessment has been a part of the educational system for centuries. The absence of theoretical, conceptual and philosophical debates with regard to assessment may, however, result in practices that tend to reproduce themselves in a vacuum, resist change, and are disconnected from relevant issues of knowledge, power and social organisation in general. (113)

Contrary to Delandshere's assertion, assessment theories do appear to have developed throughout the years. Evidence of this are the numerous academic journals whose readership have engaged with the philosophical foundations of assessment. However, the effect of these discussions on teachers' daily practice remains elusive, mainly due to the literature on teacher professional development being dominated by an emphasis on changing classroom practices (Koh 2011). Moreover, preparation of teachers in order to assess student learning effectively remains inadequate (McGee and Colby 2014). For example, Barnes (1997) reported that only 38% of the teaching staff studied had received any formal training on assessment.

Identification of differences between faculties generally stems from the work of Biglan. He identified three dimensions of academics' perceived characteristics of academic subject matter: hard or soft (the degree to which paradigm consensus exists in the field), pure or applied (the degree of concern with application of disciplinary knowledge), and life or non-life systems (1973). Subsequently, this model has been used to explore socialisation function within faculties (Creswell and Bean 1981), faculty member's intent to leave (Ryan, Healy, and Sullivan 2012), the influence of teacher's gender on female students' academic performance (Johnson 2014), and obstacles to research collaboration due to paradigmatic divergences between faculty from the hard and soft sciences (Gardner 2012).

The seminal works demonstrating differential grading standards among faculties (Aiken 1963; Goldman et al. 1974; Strenta and Elliott 1987) have made a solid foundation in the literature. More specifically, these differential grading differences have been attributed to faculty characteristics, the subject field (Prather and Smith 1976), teachers not following 'recommended grading practices for grading' (Brookhart 1993, 123), or teacher grading styles (DeBoer, Anderson, and Elfessi 2007). Teachers also view assessment as protecting the profession through gatekeeping (Nagy 2000). Additionally, Hewitt and Jacobs (1978) have shown that 'students are aware of differential grading standards in different major fields' (214).

Barnes et al. (2001) investigated faculty beliefs about grading. Significantly different beliefs about grading were revealed using the Faculty Beliefs about Grades Inventory (Barnes 1997). These differences were evident across Biglan's dimensions. However, Barnes et al. conclude that there is an indirect and direct effect of both academic discipline as well as the teaching goals of that faculty. This can be attributed to students learning the habits and traditions of a discipline and then embodying them as faculty members.

Faculties or majors that generally attract academically stronger students would have an 'average' student who would be academically stronger than those faculties that generally attract academically weaker students. Aiken (1963) reported that when admission standards were raised for one faculty, the implicit grading standards also shifted, thus leading to no significant changes in average grades for those academically stronger students.

Besides differences in faculty members' beliefs about assessment and their subsequent effects on grading, differential grading practices may be due to teachers'

Table 1. Descriptive statistics per faculty.

| | *n* | Min | $\bar{x}$ | Max | SD |
|---|---|---|---|---|---|
| Business | 910 | 4.97 | 6.52 | 8.50 | .56 |
| Industrial engineering | 169 | 5.58 | 6.62 | 8.24 | .51 |
| Education | 99 | 5.50 | 6.72 | 7.70 | .43 |
| Civil engineering | 133 | 5.50 | 6.31 | 7.94 | .39 |
| Financial management | 555 | 5.04 | 6.06 | 7.92 | .34 |
| Fine arts | 214 | 5.69 | 6.46 | 8.00 | .33 |
| Nursing | 270 | 5.39 | 6.19 | 7.60 | .34 |
| International business | 102 | 3.63 | 6.59 | 7.48 | .48 |
| Law | 58 | 5.69 | 6.33 | 7.03 | .34 |
| Social work | 362 | 5.38 | 6.39 | 7.22 | .33 |
| Environmental sciences | 208 | 5.50 | 6.12 | 9.70 | .53 |

grading practices becoming adapted to the level of the students. This is in line with Helson's (1947) assertion that the meaning of a judgement can only be understood when taking the context of that judgement into account. For example, if a teacher judges a student's essay as excellent, one can only interpret the meaning of that grade in that particular context.

Adaptation-level theory has also been investigated in the grading of essay questions (Hales and Tokar 1975), and recently the same phenomenon has been identified as the assimilation effect, also found in teacher's grading open questions (Attali 2011). In explaining adaptation-level theory, Helson states that 'fundamental to the theory is the assumption that effects of stimulation form a spatio-temporal configuration in which order prevails' (1947, 2). This order, for teachers and thus assessors, is the accumulated past experiences of all their students' performances in participating in similar classes.

However, this effect can also been seen at the micro level: the sequence in which teachers grade open questions has been demonstrated to advantage lower scoring students when their work is graded after a higher scoring open question. Attali grouped students by ethnicity and scoring level and asserts:

> because most group members will be rated immediately after a non-group member, whose score is more likely to be higher than the score of a group member. Therefore, the ratings of most group members will assimilate higher ratings that will tend to increase their scores. (2011, 75)

The current study focuses on two research questions. Firstly, is there evidence of differential grading standards among the 11 faculties in this data-set; and, secondly, is there evidence of adaptation-level in teacher's grading practices? In order to test for potential differential grading standards and adaptation-level, this study will examine the relationship of pre-entry secondary school grades (English, Dutch and Mathematics) to first-year average grades in 11 faculties. Akin to the procedure as outlined in Goldman et al. (1974), regression weights will be calculated for each student's pre-entry English, Dutch and Mathematics on average grade per faculty. These regression weights will then be used to calculate students' projected grades in each of the faculties. This will result in two sets of average grades: projected *to* (the average projected grades for all students *to* a particular faculty) and projected *from* (the average projected grades (in all majors) *from* a particular faculty). The differences between the average projected grade *to* a faculty and the averaged projected grade

Table 2. Pre-entry secondary school grades and regression equation weights per major.

| | $n$ | $r$ | English $\bar{x}$ | Dutch $\bar{x}$ | Math $\bar{x}$ | Regression equation $\bar{x}$ |
|---|---|---|---|---|---|---|
| Business | 910 | .27* | 6.30 | 6.47 | 6.65 | 4.793 + .043 (ENG) + .088 (NL) + .133 (MATH) |
| Industrial engineering | 169 | .22* | 6.75 | 6.30 | 6.78 | 5.600 + .120 (ENG) + .014 (NL) + .018 (MATH) |
| Education | 99 | .25 | 6.16 | 6.68 | 6.80 | 5.704 + .057 (ENG) + .123 (NL) + −.023 (MATH) |
| Civil engineering | 133 | .21 | 6.04 | 6.24 | 6.70 | 5.420 + −.001 (ENG) + .102 (NL) + .039 (MATH) |
| Financial management | 555 | .32* | 6.11 | 6.44 | 6.97 | 5.013 + −.033 (ENG) + .108 (NL) + .079 (MATH) |
| Fine arts | 214 | .22* | 6.91 | 6.55 | 6.43 | 5.594 + .026 (ENG) + .085 (NL) + .020 (MATH) |
| Nursing | 270 | .23* | 6.20 | 6.57 | 6.77 | 5.253 + .023 (ENG) + .106 (NL) + .013 (MATH) |
| International business | 102 | .16 | 6.98 | 6.51 | 6.73 | 6.320 + .036 (ENG) + −.061 (NL) + .062 (MATH) |
| Law | 58 | .43* | 6.55 | 6.72 | 6.53 | 6.130 + .025 (ENG) + −.126 (NL) + .136 (MATH) |
| Social work | 362 | .30* | 6.27 | 6.58 | 6.47 | 5.254 + .008 (ENG) + .072 (NL) + .094 (MATH) |
| Environmental sciences | 208 | .18 | 6.50 | 6.46 | 6.57 | 5.473 + .018 (ENG) + −.028 (NL) + .104 (MATH) |

*p<.05.

Table 3. Projected average grades from and to each major for students earning an average grade ≤6.49.

| | | Projected to | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Projected from | $n$ | Business $\bar{x}$ | Industrial engineering $\bar{x}$ | Education $\bar{x}$ | Civil engineering $\bar{x}$ | Financial management $\bar{x}$ | Fine arts $\bar{x}$ | Nursing $\bar{x}$ | International business $\bar{x}$ | Law $\bar{x}$ | Social work $\bar{x}$ | Environmental sciences $\bar{x}$ | Means $\bar{x}$ |
| Business | 450 | 6.49 | 5.95 | 6.70 | 6.32 | 6.01 | 6.43 | 6.16 | 6.56 | 6.36 | 6.37 | 6.08 | 6.31 |
| Industrial engineering | 76 | 6.54 | 6.00 | 6.70 | 6.33 | 6.02 | 6.44 | 6.17 | 6.59 | 6.42 | 6.40 | 6.12 | 6.34 |
| Education | 26 | 6.53 | 6.01 | 6.70 | 6.35 | 6.06 | 6.44 | 6.18 | 6.55 | 6.37 | 6.41 | 6.10 | 6.33 |
| Civil engineering | 96 | 6.47 | 5.92 | 6.65 | 6.30 | 6.00 | 6.41 | 6.13 | 6.57 | 6.41 | 6.37 | 6.10 | 6.30 |
| Financial management | 498 | 6.54 | 6.01 | 6.68 | 6.34 | 6.05 | 6.44 | 6.16 | 6.58 | 6.42 | 6.42 | 6.12 | 6.34 |
| Fine arts | 123 | 6.49 | 5.94 | 6.74 | 6.31 | 5.98 | 6.44 | 6.17 | 6.56 | 6.35 | 6.36 | 6.07 | 6.31 |
| Nursing | 217 | 6.53 | 6.00 | 6.70 | 6.34 | 6.05 | 6.44 | 6.17 | 6.56 | 6.38 | 6.41 | 6.10 | 6.33 |
| International business | 44 | 6.55 | 6.02 | 6.75 | 6.34 | 6.03 | 6.46 | 6.19 | 6.58 | 6.38 | 6.41 | 6.11 | 6.34 |
| Law | 40 | 6.51 | 5.98 | 6.77 | 6.35 | 6.03 | 6.47 | 6.20 | 6.53 | 6.30 | 6.39 | 6.06 | 6.32 |
| Social work | 243 | 6.48 | 5.95 | 6.71 | 6.33 | 6.01 | 6.44 | 6.17 | 6.54 | 6.33 | 6.37 | 6.07 | 6.31 |
| Environmental sciences | 172 | 6.50 | 5.96 | 6.72 | 6.32 | 6.01 | 6.44 | 6.17 | 6.56 | 6.36 | 6.38 | 6.09 | 6.31 |
| Means | | 6.51 | 5.98 | 6.71 | 6.33 | 6.02 | 6.44 | 6.17 | 6.56 | 6.37 | 6.39 | 6.09 | |

Table 4. Effect sizes for projected to and projected from for students earning an average grade ≤6.49.

| | Projected to | | | Projected from | | | |
|---|---|---|---|---|---|---|---|
| Major | $\bar{x}$ | SD | Var. | $\bar{x}$ | SD | Var. | Cohen's $\delta$ |
| Business | 6.51 | 0.16 | 0.02 | 6.31 | 0.10 | 0.01 | 1.50 |
| Industrial engineering | 5.98 | 0.19 | 0.04 | 6.34 | 0.10 | 0.01 | −2.31 |
| Education | 6.71 | 0.10 | 0.01 | 6.34 | 0.11 | 0.02 | 3.32 |
| Civil engineering | 6.33 | 0.08 | 0.01 | 6.30 | 0.11 | 0.01 | 0.28 |
| Financial management | 6.02 | 0.11 | 0.01 | 6.34 | 0.11 | 0.01 | −2.83 |
| Fine arts | 6.44 | 0.07 | 0.00 | 6.31 | 0.11 | 0.01 | 1.38 |
| Nursing | 6.17 | 0.08 | 0.01 | 6.33 | 0.11 | 0.01 | −1.65 |
| International business | 6.56 | 0.07 | 0.00 | 6.35 | 0.11 | 0.01 | 2.25 |
| Law | 6.37 | 0.14 | 0.02 | 6.32 | 0.11 | 0.01 | 0.36 |
| Social work | 6.39 | 0.11 | 0.01 | 6.31 | 0.10 | 0.01 | 0.75 |
| Environmental sciences | 6.09 | 0.10 | 0.01 | 6.32 | 0.12 | 0.02 | −2.00 |

*from* a faculty will be compared. If differential grading standards due to adaptation-level exist, the data should reveal that majors with academically stronger students grade more stringently and majors with academically weaker students grade less stringently. This will be investigated by correlation analysis and calculating standardised effect size measures (Cohen's $\delta$) between *projected to* and *projected from* average scores.

## Participants

A data-set was extracted from the central administration containing students who were successful in earning their 'first year' certificate during the first year of studies at a Dutch university of applied sciences ($n = 3080$). In Dutch higher education, students are generally afforded two years to complete this certificate. For this study, these students were selected in order to create a sample that did not include other types of students, such as those who partially followed the programme and then subsequently dropped out, or those who were still actively pursuing their first-year certificate. This choice was mainly based on the availability of data: students who dropped out still have a year average grade, but only for a few subjects, and students still pursuing their year-one certificate have been afforded re-examinations that might potentially influence their average grade. Additionally, a sub sample of students who followed higher general secondary education [Hoger algemeen voortgezet onderwijs] ($n = 2229$) or pre-university secondary education [Voorbereidend wetenschappelijk onderwijs] ($n = 848$) was used. These two types of students form approximately 80% of all incoming first-year students, and their secondary school grades are directly imported into the central administration database upon the student registering for a programme (see Tables 1 and 2).

## Results

Analogous to the work of Goldman et al. (1974), two groups of students were created in the data using the earned average grade in the first year: average ability (<6.50) and high ability (≥6.5). The Dutch education system uses a 1 to 10-point scale for assessments, where 6.5 *roughly* approximates to a 'B' in the American grading system.

Table 5. Projected average grades from and to each major for students earning an average grade ≥6.5.

| Projected from | *n* | Business $\bar{x}$ | Industrial engineering $\bar{x}$ | Education $\bar{x}$ | Civil engineering $\bar{x}$ | Projected to Financial management $\bar{x}$ | Fine arts $\bar{x}$ | Nursing $\bar{x}$ | International business $\bar{x}$ | Law $\bar{x}$ | Social work $\bar{x}$ | Environmental sciences $\bar{x}$ | Means $\bar{x}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Business | 460 | 6.54 | 6.02 | 6.71 | 6.34 | 6.05 | 6.45 | 6.18 | 6.57 | 6.39 | 6.41 | 6.11 | 6.34 |
| Industrial engineering | 93 | 6.54 | 5.98 | 6.71 | 6.31 | 5.99 | 6.44 | 6.16 | 6.60 | 6.43 | 6.39 | 6.12 | 6.33 |
| Education | 73 | 6.56 | 6.04 | 6.73 | 6.36 | 6.07 | 6.46 | 6.20 | 6.56 | 6.37 | 6.43 | 6.11 | 6.35 |
| Civil engineering | 37 | 6.55 | 6.02 | 6.69 | 6.35 | 6.06 | 6.44 | 6.17 | 6.57 | 6.41 | 6.42 | 6.12 | 6.34 |
| Financial management | 57 | 6.62 | 6.12 | 6.70 | 6.38 | 6.12 | 6.47 | 6.20 | 6.59 | 6.44 | 6.48 | 6.16 | 6.38 |
| Fine arts | 91 | 6.56 | 6.03 | 6.78 | 6.35 | 6.03 | 6.48 | 6.21 | 6.57 | 6.36 | 6.41 | 6.10 | 6.35 |
| Nursing | 53 | 6.59 | 6.08 | 6.76 | 6.38 | 6.07 | 6.48 | 6.22 | 6.56 | 6.36 | 6.44 | 6.11 | 6.36 |
| International business | 58 | 6.57 | 6.02 | 6.75 | 6.34 | 6.01 | 6.46 | 6.19 | 6.60 | 6.41 | 6.41 | 6.12 | 6.35 |
| Law | 18 | 6.60 | 6.08 | 6.73 | 6.36 | 6.07 | 6.47 | 6.20 | 6.59 | 6.42 | 6.45 | 6.14 | 6.37 |
| Social work | 119 | 6.54 | 6.02 | 6.74 | 6.36 | 6.05 | 6.46 | 6.20 | 6.55 | 6.34 | 6.41 | 6.09 | 6.34 |
| Environmental sciences | 36 | 6.57 | 6.04 | 6.73 | 6.35 | 6.04 | 6.46 | 6.19 | 6.59 | 6.41 | 6.43 | 6.13 | 6.35 |
| Means | | 6.57 | 6.04 | 6.73 | 6.35 | 6.05 | 6.46 | 6.19 | 6.58 | 6.39 | 6.43 | 6.12 | |

Table 6. Effect sizes for projected to and projected from for students earning an average grade ≥6.5.

| Major | Projected to | | | Projected from | | | Cohen's $\delta$ |
|---|---|---|---|---|---|---|---|
| | $\bar{x}$ | SD | Var. | $\bar{x}$ | SD | Var. | |
| Business | 6.57 | 0.18 | 0.03 | 6.34 | 0.11 | 0.01 | 1.49 |
| Industrial engineering | 6.04 | 0.22 | 0.05 | 6.34 | 0.13 | 0.02 | −1.61 |
| Education | 6.73 | 0.11 | 0.01 | 6.35 | 0.11 | 0.01 | 3.29 |
| Civil engineering | 6.35 | 0.09 | 0.01 | 6.34 | 0.13 | 0.02 | 0.08 |
| Financial management | 6.05 | 0.12 | 0.01 | 6.39 | 0.12 | 0.02 | −2.71 |
| Fine arts | 6.46 | 0.08 | 0.01 | 6.35 | 0.12 | 0.02 | 1.02 |
| Nursing | 6.19 | 0.09 | 0.01 | 6.37 | 0.11 | 0.01 | −1.70 |
| International business | 6.58 | 0.07 | 0.01 | 6.35 | 0.13 | 0.02 | 2.06 |
| Law | 6.39 | 0.14 | 0.02 | 6.37 | 0.13 | 0.02 | 0.16 |
| Social work | 6.43 | 0.12 | 0.01 | 6.34 | 0.11 | 0.01 | 0.70 |
| Environmental sciences | 6.12 | 0.11 | 0.01 | 6.36 | 0.12 | 0.02 | −2.05 |

Table 3 presents the projected *to* and *from* average grades for students earning an average grade ≤6.49 in their first year of studies. For example, the first row presents the projected average grades from business students to the other majors. On average, using the regression weights, the projected business student with their pre-entry level would earn a 6.70 in education. The means at the right of the table are for all the projected from majors. The second column is the projected *to* means for all the majors to business. On average, all students would earn a projected mean of 6.51 to business. When investigating the relationship between the projected to and projected from tests, a negative correlation ($r = -.15$) was revealed. This means that majors whose students were projected to earn the highest scores would expect to award the lowest grades to students from other fields. Financial management students have the highest projected from mean ($m = 6.342$) and the lowest projected to means ($m = 6.02$)

Differences in means between *projected to* and *projected from* scores were analysed by calculating standardised effect size measures (Cohen's $\delta$). Cohen (1988) defines a 'small' effect as 0.2, a medium effect as 0.5 and large as 0.8. For students earning an average grade ≤6.49, financial management has the largest negative effect ($\delta = -2.83$), and education has the largest positive effect ($\delta = 3.32$). This means that, on average, assessors from financial management graded two standard deviations more stringently. Conversely, education assessors grade more than three standard deviations less stringently (see Table 4).

For students earning an average grade ≤ 6.49, a negative correlation was found between projected to and projected from mean scores ($r = -.32$). For higher achieving students, financial management majors had the highest projected from mean scores ($m = 6.389$), while the assessors from business informatics graded the most stringently ($m = 6.04$). Education assessors graded the least stringently ($m = 6.73$) (see Table 5).

Tests for means differences between *projected to* and *projected from* scores were conducted (see Table 6), and revealed that financial management had the largest negative effect ($\delta = -2.71$). Education had largest positive effect ($\delta = 3.29$).

## Conclusion

The results presented in this study demonstrate the presence of differential grading standards among the 11 faculties within this data-set. This is evidenced by the

significant differences found between *projected to* and *projected from* scores. In other words, if a student followed another major, there is a high chance that they would earn a significantly different grade in that new major.

Besides the presence of differential grading standards, this research also confirms the implications of adaptation-level theory in the grading standards among the different faculties. The significant correlations between *projected to* and *projected from* scores (both negative and positive) are consistent with adaptation-level theory. Faculties that attracted academically weaker students generally grade less stringently, and thus these students received on average lower grades when projected from these weaker faculties. Conversely, faculties that attracted academically stronger students generally grade more stringently, and thus these students received on average higher grades when projected from these stronger faculties

The explained variance for the regression of pre-entry secondary school grades on earned average grade differs per faculty. These differences in grading have been generally attributed to either the varying nature of the tasks needing to be performed successfully within that faculty, or the differences among 'performance' majors (such as fine arts) (Goldman and Hewitt 1975). However, this argument is not supported by this data, due to the largest positive effect size being for education, followed by international business, which would not generally be characterised as 'performance' majors.

## Implications for practice

Focus on teachers' assessment literacy has shown the positive value of professional development on assessment practices (Koh 2011; McGee and Colby 2014; Plake 1993). However, a gap between academic knowledge and practice remains (Popham 2009). Additionally, Popham raises the question as to whether or not assessment literacy is a real focus or just a fad. He emphatically asserts that assessment literacy is not a fad, and outlines 13 target skills and knowledge areas essential for professional development programmes. He argues that:

> until preservice teacher educators routinely provide meaningful assessment literacy for prospective teachers, the architects of professional development programs will need to offer assessment-literacy programs. (11)

The notion that students, with similar abilities when starting their academic career, can earn statistically different grades based in part on their faculty should be a call to all educators to re-examine grading practices. Of course, there are many external variables in determining grades to be found in student and institutional characteristics. However, that faculties which attract academically stronger students generally grade more stringently remains problematic. This could partly be explained by teachers acting as gatekeepers in order to protect the profession and their reputation. Equally problematic are the grading practices of teachers from faculties that attract academically weaker students. Their less stringent grading could be seen as grade inflation. This being said, the combination of these two problematic situations creates an environment of unfairness and inequality: disadvantageous to those academically stronger students who receive lower grades, and disingenuous to those academically weaker students whose grades do not form an honest assessment of their work.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Notes on contributor

Brian P. Godor is an assistant professor at the Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands. His research interests include teacher competencies concerning assessment and classroom differentiation specifically for gifted students. His recent publications appear in Assessment and Evaluation in Higher Education and Teaching in Higher Education.

## ORCID


*Brian P. Godor* http://orcid.org/0000-0003-3070-3404

## References

Aiken, L. R. 1963. "The Grading Behavior of a College Faculty." *Educational and Psychological Measurement* 23 (2): 319–322. http://doi.org/10.1177/001316446302300209.

Attali, Y. 2011. "Sequential Effects in Essay Ratings." *Educational and Psychological Measurement* 71 (1): 68–79. http://doi.org/10.1177/0013164410387344.

Barnes, L. 1997. "Development of the Faculty Beliefs about Grades Inventory." *Educational and Psychological Measurement* 57 (3): 459–468. http://doi.org/10.1177/0013164497057003007.

Barnes, L., K. S. Bull, N. J. Campbell, and K. M. Perry. 2001. "Effects of Academic Discipline and Teaching Goals in Predicting Grading Beliefs among Undergraduate Teaching Faculty." *Research in Higher Education* 42 (4): 455–467.

Biglan, A. 1973. "The Characteristics of Subject Matter in Different Academic Areas." *Journal of Applied Psychology* 57: 195–203.

Boud, D. 1990. "Assessment and the Promotion of Academic Values." *Studies in Higher Education* 15 (1): 101–111. http://doi.org/10.1080/03075079012331377621.

Brookhart, S. M. 1993. "Teachers' Grading Practices: Meaning and Values." *Journal of Educational Measurement* 30 (2): 123–142.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates.

Creswell, J. W., and J. P. Bean. 1981. "Research Output, Socialization, and the Biglan Model." *Research in Higher Education* 15 (1): 69–91. http://doi.org/10.1007/BF00976549.

DeBoer, B. V., D. M. Anderson, and A. M. Elfessi. 2007. "Grading Styles and Instructor Attitudes." *College Teaching* 55 (2): 57–64. http://doi.org/10.3200/CTCH.55.2.57-64.

Delandshere, G. 2001. "Implicit Theories, Unexamined Assumptions and the Status Quo of Educational Assessment." *Assessment in Education: Principles, Policy & Practice* 8 (2): 113–133. http://doi.org/10.1080/09695940123828.

Gardner, S. K. 2012. "Paradigmatic Differences, Power, and Status: A Qualitative Investigation of Faculty in One Interdisciplinary Research Collaboration on Sustainability Science." *Sustainability Science* 8 (2): 241–252. http://doi.org/10.1007/s11625-012-0182-4.

Gibbs, G. 2010. *Using Assessment to Support Student Learning*. Leeds: Leeds Metropolitan University.

Goldman, R. D., and B. N. Hewitt. 1975. "Adaptation-Level as an Explanation for Differential Standards in College Grading." *Journal of Educational Measurement* 12 (3): 149–161.

Goldman, R. D., D. E. Schmidt, B. N. Hewitt, and R. Fisher. 1974. "Grading Practices in Different Major Fields." *American Educational Research Journal* 11 (4): 343–357. http://doi.org/10.3102/00028312011004343.

Hales, L. W., and E. Tokar. 1975. "The Effect of the Quality of Preceding Responses on the Grades Assigned to Subsequent Responses to an Essay Question." *Journal of Educational Measurement* 115–117.
Helson, H. 1947. "Adaptation-Level as Frame of Reference for Prediction of Psychophysical Data." *The American Journal of Psychology* 60 (1): 1. http://doi.org/10.2307/1417326
Hewitt, B. N., and R. Jacobs. 1978. "Student Perceptions of Grading Practices in Different Major Fields." *Journal of Educational Measurement* 15 (3): 213–218.
Hills, J. R. 1991. "Apathy concerning Grading and Testing." *The Phi Delta Kappan* 72 (7): 540–545.
Johnson, I. Y. 2014. "Female Faculty Role Models and Student Outcomes: A Caveat about Aggregation." *Research in Higher Education* 55 (7): 686–709. http://doi.org/10.1007/s11162-014-9331-1.
Koh, K. H. 2011. "Improving Teachers' Assessment Literacy through Professional Development." *Teaching Education* 22 (3): 255–276. http://doi.org/10.1080/10476210.2011.593164
McGee, J., and S. Colby. 2014. "Impact of an Assessment Course on Teacher Candidates' Assessment Literacy." *Action in Teacher Education* 36 (5–6): 522–532. http://doi.org/10.1080/01626620.2014.977753.
Nagy, P. 2000. "The Three Roles of Assessment: Gatekeeping, Accountability, and Instructional Diagnosis." *Canadian Journal of Education/Revue Canadienne De L'éducation* 25 (4): 262–279. http://doi.org/10.2307/1585850
Plake, B. S. 1993. "Teacher Assessment Literacy: Teachers' Competencies in the Educational Assessment of Students." *Midwestern Educational Researcher* 6 (1): 21–27.
Popham, W. J. 2009. "Assessment Literacy for Teachers: Faddish or Fundamental?" *Theory into Practice* 48 (1): 4–11. http://doi.org/10.1080/00405840802577536.
Prather, J. E., and G. Smith. 1976. "A Study of the Relationships between Faculty Characteristics, Subject Field, and Course Grading Patterns." *Research in Higher Education* 5 (4): 351–363. http://doi.org/10.1007/BF00993434.
Ryan, J. F., R. Healy, and J. Sullivan. 2012. "Oh, Won't You Stay? Predictors of Faculty Intent to Leave a Public Research University." *Higher Education* 63 (4): 421–437.
Strenta, A. C., and R. Elliott. 1987. "Differential Grading Standards Revisited." *Journal of Educational Measurement* 24 (4): 281–291.