

# Predicting different grades in different ways for selective admission: disentangling the first-year grade point average

Sebastiaan C. Steenman, Wieger E. Bakker & Jan W.F. van Tartwijk

To cite this article: Sebastiaan C. Steenman, Wieger E. Bakker & Jan W.F. van Tartwijk (2016) Predicting different grades in different ways for selective admission: disentangling the first-year grade point average, *Studies in Higher Education*, 41:8, 1408-1423, DOI: 10.1080/03075079.2014.970631

To link to this article: <http://dx.doi.org/10.1080/03075079.2014.970631>



Published online: 11 Nov 2014.



Submit your article to this journal [↗](#)



Article views: 481



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

## Predicting different grades in different ways for selective admission: disentangling the first-year grade point average

Sebastiaan C. Steenman<sup>a\*</sup>, Wieger E. Bakker<sup>a</sup> and Jan W.F. van Tartwijk<sup>b</sup>

*<sup>a</sup>School of Governance, Utrecht University, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands; <sup>b</sup>Faculty of Social and Behavioral Sciences, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands*

The first-year grade point average (FYGPA) is the predominant measure of student success in most studies on university admission. Previous cognitive achievements measured with high school grades or standardized tests have been found to be the strongest predictors of FYGPA. For this reason, standardized tests measuring cognitive achievement are widely used as a tool for selective admission to higher education. The FYGPA, however, measures many markedly different aspects of student success. In this article it is shown that when the FYGPA is divided into averages that reflect performance on different types of goals, the predictive value of previous cognitive achievement differs significantly between these disentangled averages. It is therefore important to distinguish between different types of goals when considering what student success is, and which students should be admitted to particular university programmes.

**Keywords:** selective admissions; student success; grades; taxonomy of educational objectives; Dublin descriptors

### Introduction

Admitting students to universities solely on the basis of their previous cognitive performance, measured with standardized tests (predominantly either the SAT or the ACT) or high school grades, is generally met with resistance from educators and university administrators. An important reason for this resistance is that educators feel that academic success should be more than mere cognitive performance (cf. Soares 2012). However, research again and again shows that such measures of previous cognitive performance are, by far, the best predictors of academic success (Bhattacharya, Kanaya, and Stevens 2012; Ferguson, James, and Madeley 2002; Friedman and Mandel 2009; Gifford, Briceño-Perriott, and Mianzo 2006, 23; Harackiewicz et al. 2002; Kuncel and Hezlett 2010; McKenzie, Gow, and Schweitzer 2004; Reumer and Van der Wende 2010; Robbins et al. 2004, 273; Sackett, Borneman, and Connelly 2008; Siu and Reiter 2009). These empirical findings can be explained by the undifferentiated way in which student success is usually measured, that is, with first-year grade point average (FYGPA).

Several researchers have differentiated between measures for student success already. Stemler (2012), for example, explores the multiplicity of goals that universities

---

\*Corresponding author. Email: [s.c.steenman@uu.nl](mailto:s.c.steenman@uu.nl)

set for students in addition to achieving high average grades, and Schultz and Zedeck (2012) introduce a focus on career success instead of focussing on academic success of students at university. Still, most academic work on academic success continues to define success (explicitly or implicitly) as a high FYGPA (Bhattacharya, Kanaya, and Stevens 2012; Bridgeman, McCamley-Jenkins, and Ervin 2000; De Vries, De Vries, and Born 2011; Friedman and Mandel 2009; Gifford, Briceño-Perriott, and Mianzo 2006; Komarraju, Karau, and Schmeck 2009; Petska 2006; Richardson, Abraham, and Bond 2012; Uppal and Mishra 2013; Urlings-Strop et al. 2009; Te Wierik, Bieshuizen, and Van Os 2014; Wintre et al. 2011; see also Stemler 2012).

This is a logical approach because of the lack of other reliable and valid indicators for student success. The development of leadership skills, for instance, is a goal of many degree programmes (Schmitt 2012), but is difficult to measure and is therefore less suitable as an additional indicator for success. Measuring these types of goal achievement (and the potential to do so) seems to have to rely on self-reported measures that might work somewhat differently when potential students are aware of what the stakes are in these tests before admission (see, for example, the measures developed by Schmitt [2012]). Whether former students have successful careers is not a good alternative, because it is extremely complicated to measure as well. This is especially the case when career paths become more diffuse and when a focus on level of income or self-reported success is unsatisfactory. Where the success of lawyers after law school might be compared (see the very nuanced and interesting approach by Schultz and Zedeck 2012), the career success of, say, former sociology students is likely to be much harder to measure uniformly.

In this article we will explore an alternative approach: differentiating in the grades that constitute the FYGPA for the different types of goals that are to be attained. We will link these goals to the 'Dublin descriptors', which describe which elements need to be part of an academic degree programme in countries that take part in the Bologna process.

### *Different types of goals in the cognitive domain*

The influential taxonomy of educational objectives by Bloom (1956) and his colleagues has spurred a large body of research into the differentiation of learning goals. It also states explicitly that these differing goals should be measured (assessed) in different ways (Bloom 1956, 3). For instance, the lowest level of cognitive goals, knowledge of terminology, will be assessed using questions asking for definitions of technical terms (in either short-answer or multiple-choice form), while the highest level, judgments in terms of external criteria, will likely entail some sort of written assignment about, for example, the comparison of major theories and facts about a certain theoretical concept (Bloom 1956, 201–7). In the revision of this taxonomy, edited by Anderson and Krathwohl (2001), this assumption that different types of assessments are needed for different levels of the cognitive domain is reiterated.

Both the original and the revised taxonomies state that the different parts of the cognitive domain are in a hierarchical relation to each other on a scale from simple to complex. On first glance, this might mean that the correlation between grades for the different levels in the cognitive domain and previous cognitive performance would be similar. When only the level of difficulty differs, all levels should be predictable in a comparable way, but performance on these 'higher levels' of cognitive performance would only be somewhat lower than performance at the 'lower levels' within

the cognitive domain. But actually, the connection between the levels is somewhat different. When, for example, looking at the step between *understanding* (or *comprehension*) and *applying*, it seems rather logical that understanding something is a necessary basis for applying it. Lower cognitive abilities are needed to attain a good performance on the higher levels in the cognitive domain, but this does not, however, mean that the higher skills are simply more of the same. *Applying* adds a distinctive and separate element to *understanding*. Therefore, it is a logical conclusion that previous performance on the *understanding* level of the cognitive domain still correlates with performance on the higher levels of that domain, but to a lesser and lesser degree when higher cognitive abilities are measured. Elements other than understanding, such as the ability to apply the understood knowledge to specific situations or the ability to create new insights based on that understanding, are part of the measurement in these levels as well. The meta-analysis by Anderson and Krathwohl (2001, 289–91) supports this assumption about the correlations between the different levels of the cognitive domain.

So, it is likely that different levels of the cognitive domain are part of the assessment programme of universities, and that these levels are tested in different ways. These tests reflect, at least partly, different *types* of cognitive abilities (lower vs. higher order; from the level *remember* to the level *create*).

### ***Academic performance in the non-cognitive domains***

It is likely that not only different levels of cognitive performance will be part of the FYGPA statistic, but also performance in the other domains that Bloom (1956) and his colleagues distinguish. Elements of what might be called *academic skills* (including aspects such as critical evaluation, academic ethics, and being able to present insights in written and spoken word) are part of the 'affective domain' in Bloom's traditional taxonomy. Classroom behaviour and communication, as well as normative evaluation of perceived behaviour and external events, seem to clearly belong in the affective domain (see Krathwohl, Bloom, and Masia 1964). In their description of the renewed taxonomy of the cognitive domain, Anderson and Krathwohl (2001, 258–59) acknowledge that affective skills play a role next to elements from the cognitive domain. Furthermore, in a number of academic degree programmes psychomotor skills are important as well, probably most prominently in medical and technical fields.

There are clear indications that testing performance on different domains happens in different ways. Anderson and Krathwohl (2001, 60–61) show this in their discussion of the testing of meta-cognitive knowledge, about which they argue that there is a large overlap with the affective domain. They argue that in this dimension, largely due to this overlap with the affective domain, it is much less likely to find one correct answer. They argue that this is due to individual characteristics in learning processes, which ask for a different type of assessment, more focussed on how individual students have internalized a system of communication and valuing. The original taxonomy of the affective domain (Krathwohl, Bloom, and Masia 1964) takes a similar approach, also pointing towards different forms of assessment for different levels of the affective domain, and certainly different from the assessment of skills that are in the cognitive domain. In predicting future performance, it is therefore likely that previous cognitive performance can predict performance on affective skills to a much smaller degree.

### *Dublin descriptors*

That these different types of goals, both in the cognitive and in the affective domain, play a role in many university programmes is clearly visible in the Dublin descriptors. These descriptors were developed as part of the Bologna process in which European countries aimed to develop ‘a coherent and cohesive European Higher Education Area’ (Berlin Communiqué 2003, 1). Not only has this resulted in a more coherent system of bachelor and master degrees between many European countries, but part of the process is also to ‘elaborate a framework of comparable and compatible qualifications ... which should seek to describe qualifications in terms of workload, level, learning outcomes, competencies and profile’ (Berlin Communiqué 2003, 4).

The result of this process is a framework with five different core goals that should be part of each degree programme, and which should be mastered at different levels depending on the degree (bachelor, master, or PhD). These goals are

- knowledge and understanding;
- applying knowledge and understanding;
- making judgements;
- communication; and
- learning skills (Joint Quality Initiative Group 2004).

By the development of these goals, the countries that take part in the Bologna process have shown that there is broad support for a multi-dimensional approach to learning outcomes in higher education. Table 1 shows which elements of the revised cognitive taxonomy and the affective taxonomy are part of each of the Dublin descriptors. The first three of the Dublin descriptors represent different levels of the cognitive domain. *Knowledge and understanding* represents the

Table 1. How Dublin descriptors relate to the cognitive and affective domain.

Dublin descriptor	Cognitive domain		Affective domain
	Cognitive process dimension	Knowledge dimension	
Knowledge and understanding	Remember Understand	Factual knowledge Conceptual knowledge Procedural knowledge	
Applying knowledge and understanding	Apply Analyse	Factual knowledge Conceptual knowledge Procedural knowledge	
Making judgement	Analyse Evaluate Create	Factual knowledge Conceptual knowledge Procedural knowledge	Valuing Organization Characterization
Communication			Receiving Sending
Learning skills	All	Meta-cognitive knowledge	Meta-cognitive elements from different levels

*remember* and *understand* levels of the revised cognitive taxonomy, while *applying knowledge and understanding* holds the *apply* level of the revised taxonomy, as well as elements of the *analyse* level. This *applying knowledge and understanding* descriptor also contains goals about more academic application in using theory to see new research paths (Joint Quality Initiative Group 2004), while *applying* in the taxonomy refers to more direct applications of single theories. The descriptor *making judgements* involves skills that are described in the *analyse*, *evaluate* as well as the *create* level of the revised cognitive taxonomy. It holds that students are able to analyse data, integrate knowledge, and judge by drawing conclusions from observed data. This descriptor also includes organizing research processes. On the other hand, *making judgements* also holds critical analysis, reflecting on social and ethical questions and responsibilities (Joint Quality Initiative Group 2004). These elements are part of the affective domain as described by Bloom (1956) and Krathwohl, Bloom, and Masia (1964), more specifically the three highest levels of that domain: *valuing*, *organization* (of an ethical and social system), and *characterization* (using the system on knowledge production to consider broader consequences).

In the last two descriptors, the focus is on skills. *Communication* is squarely in the affective domain (at the levels of both *receiving* and *sending*). The descriptor *learning skills* is closest to what Anderson and Krathwohl (2001) call *meta-cognitive* knowledge (this descriptor holds goals such as learning skills to continue learning on a higher level, more and more self-directed). As mentioned before, Anderson and Krathwohl (2001) consider the meta-cognitive type of knowledge to 'bridge the cognitive and affective domains' (259). On the other hand, testing of meta-cognitive skills is rather distinct from the testing of other cognitive elements (Anderson and Krathwohl 2001, 60). The descriptor *learning skills* seems to lean on both the cognitive and affective domains, without being a separate level in either one of them.

Even though educators seem to struggle somewhat with the implementation of the last two Dublin descriptors in their teaching and testing (Kehm 2010), it is likely that these different descriptors will be part of higher education in the countries that take part in the Bologna process. Furthermore, wherever these more distinct learning outcomes are part of the assessment programmes, different types of tests are likely to measure performance in these distinct competencies.

### ***Different measures, different predictors***

In the previous paragraphs it was argued that there are fundamental differences between what different tests are likely to measure. First, it was argued that different levels of the cognitive domain are measured in different ways, and, second, that performance in different domains is measured differently. When studying which measures predict future performance in higher education, it is therefore logical to distinguish between performance on these levels and in these domains. Different elements in the admission process will be able to predict different grades, either better or worse, depending on in which domain and on which level this element measures itself. High school grade point average (HSGPA) or standardized tests such as the SAT or ACT are likely to predict those grades best, which most resemble what they themselves measure. There are two reasons for why this is likely to be *knowledge and understanding*. The first is that in most educational systems, HSGPA will reflect the lower levels of cognitive skills better than the higher,

because in many examinations standardized tests are dominant that are better suited for the lower cognitive levels. The second reason is that secondary education is expected to provide the basis for higher education and is therefore likely to focus on lower levels of skills. In contrast, higher levels in the cognitive domain are likely to be more prominent in higher education where students specialize and build on the knowledge base they developed in secondary education. Following this logic, HSGPA will very likely be able to predict grades obtained at university focussed on assessing lower level cognitive skills better than the higher level cognitive skills. And since high school testing is mostly cognitive, it is also likely to predict future performance in the cognitive domain better than performance in the affective domain. The SAT claims to measure knowledge and the application of knowledge (College Board 2014), although claims about measuring critical thinking are also made (Atkinson and Geiser 2009). The focus of these tests is, however, predominantly on the cognitive domain, and even when some attempts might be made to measure performance on higher levels in the cognitive domain, the form of the test itself causes a focus on lower level cognitive skills (as we have also argued earlier, the type of test differs for different levels in the cognitive domain).

There are studies in which separate grades within FYGPA are studied. Jacobs (2010), for example, compares the potential of different grades to predict performance in different courses, but focusses on differences in subject matter. Other studies distinguish between performance in different courses or stages of the programme mainly to see the predictive value of high school grades through time (Urlings-Strop et al. 2011), or simply seem focussed on various high school grades' potential for predicting the results of different courses (e.g. Hadsell [2010]; see also Richardson, Abraham, and Bond [2012]). No studies have, as far as we know, theoretically or empirically distinguished between performances on different kinds of goals to see whether these differences have any bearing on predictive value of admission instruments.

Being able to substantiate the use of different measures in the admission process is of continuing, and increasing, importance. On the one hand, it substantiates that being a successful university student constitutes more than a high GPA. On the other hand, there is a need to be able to correlate other admission instruments to some desired outcome to be able to account for their use. The strongest case to make seems to be for measures that are able to predict performance on descriptors other than knowledge and understanding. Furthermore, a clear view on what student success is seems necessary to do this. Measures objectively predicting a more nuanced view of future performance can negate the problems of bias in the admission process (Dunne, King, and Ahrens 2013; Reumer and Van der Wende 2010; Soares 2012; Zimdars 2010). A specific case in point can be found in the Netherlands, where new legislation (Kwaliteit in Verscheidenheid Act of 2013) forces schools<sup>1</sup> that are allowed to have selective admission (most are not) to use at least two admission instruments to decide on who will be able to enter and who would not. Clearly, Parliament in the Netherlands felt the same unease with the use of just previous cognitive performance assessments, and has found this unacceptable since its first discussion of selective admission in 1966 (De Jong 2013).

The question in this article is therefore: Are grades at the university level that measure the performance on different types of goals equally well predicted by high school GPA as a measure of previous cognitive performance?



## Methods

### *Participants*

In this study, data from a sample of students taking part in the same programme (including being graded for the same tests) were used. This sample consisted of a cohort of Public Administration and Organisational Science (PAOS; a full first-cycle bachelor programme) students at a large Dutch university. In the Netherlands, the Dublin descriptors are the basis for the Dutch National Qualifications Framework, which is used to evaluate and certify academic programmes (QANU 2012). Therefore, Dutch universities have implemented the Dublin descriptor framework. The official evaluation of the programme by an independent committee judged the assessment methods of this particular programme to ‘exceed the requirements’ and stated, ‘It [the committee] has established that the learning outcomes correspond sufficiently to the Dublin descriptors ...’ (QANU and EAPAA 2011, 27). Furthermore, this committee states:

The methods used properly reflect the level and orientation of the programmes. Written exams, for instance, are used in the assessment of courses [that] focus on the acquisition of knowledge and insight, while practical tests are used to assess whether students have acquired professional skills ...

The committee applauds that students of [this] programme are continually assessed and are subjected to great variety of assessment methods that are highly appropriate for both the content and working method of the courses. (QANU and EAPAA 2011, 47–48)

Next to that, the PAOS programme uses selective admission, and therefore detailed data on high school performance are available. Concerning FYGPA, the dependent variable, the programme uses a rather large array of different kinds of tests and has a first year that is the same for all students who are admitted to the programme, thereby lending it especially well for the kind of analysis that this study asks for.

### *Data collection*

For this study, data were collected about the 2012 cohort of the PAOS programme, both about the admission process and the grades that were achieved after admission.

### *HSGPA – independent variable*

Data were collected to calculate HSGPA from official high school transcripts. The averages (HSGPA) were coded on the regular Dutch grading scale from 0 to 10. In effect, because of the demands of graduating high school, the range can be between 5.5 (the minimum passing grade) and 10. HSGPA and official final course grades are (as stipulated by the Dutch high school education laws) made up of the grades achieved for the central, government administered, examination (half of the grade for most subjects) and school examinations (the other half, some subjects having only school examinations). The central examinations are always written examinations with a mix of knowledge and understanding, and usually some application of that knowledge. Most of the school examinations measure these cognitive levels as well because the Dutch Inspectorate for Secondary Education checks whether the differences between the central and school examinations are not too large as a part of their quality



control system. For schools and teachers this is an incentive to use school examinations that measure ability in a similar way as the central government examinations.

Even though Cronbach's alpha for the five courses most taken by students that were included in the analyses gives evidence for a one-factor structure underlying the HSGPA (Cronbach's alpha is 0.79), data were also collected on the final grade for two specific subjects: the average for tests (both final central written examination and school examination) on the native language of students (Dutch) and the average of tests in mathematics. Since there are different types of mathematics that students in Dutch high schools can choose between, and since only one of those were taken by more than thirty students in the sample, only students taking this type (named Mathematics A, which is focussed primarily on probability calculus) were included in all analyses where high school performance in mathematics was used.

#### *University grades – dependent variable*

Since most studies seem to use FYGPA (see before), grades that were achieved in the first year of this cohort were collected from the university's central registration system. In the analysis, only grades for 'first attempts' were taken into account (i.e. results for resits were excluded) to make sure that the grades used in the analysis are as comparable as possible, and to enlarge the range of grades.

In the Dublin descriptors, three different levels of cognitive performance are distinguished: *knowledge and understanding*; *applying knowledge and understanding*; and *making judgements*. This is a substantial reduction of the levels that are distinguished in both the Bloom and the Anderson and Krathwohl taxonomies, but in distinguishing between tests as a whole, the more crude distinction made in the Dublin descriptors appears to be more reliable. Grades that predominantly measure communication or learning skills (though less numerable) were categorized as well. This categorization was done by examining the tests themselves as well as the education goals that they were meant to assess. The focus in the categorization was on what the grades for the specific tests most prominently represent. Whenever there was any doubt, the examiner of the course was consulted. It is important to note that the categorization is based on the most prominent descriptor being measured. In some cases (mainly written examinations) there seems to be some overlap between knowledge and understanding and their application on cases. These tests were, however, usually focussed on assessing the understanding of theory, rather than on independent application.

To test the reliability of the categorization, a second coder was asked to make the categorization. This second coder was not in any other way directly involved in this study, but did have a thorough understanding of the first-year curriculum of the PAOS programme. This coder received information about the Dublin descriptors as presented in the document by the Joint Quality Initiative Group (2004), as well as about the linking between the Dublin descriptors to the taxonomy as provided earlier, but no information about the categorization made by the first coder (the first author of this article). Subsequently, the inter-coder reliability was calculated, showing a Cohen's kappa of 0.78. After the initial coding by both coders, the differences between them were discussed. This discussion showed that the differences in coding all concerned differences between the three descriptors that primarily contain levels in the cognitive domain. It was furthermore concluded that these differences are the result of different conclusions about the question what the most prominent descriptor in the grading was. Because these tests did not clearly enough measure only one of the descriptors, and since

comparing between the descriptors is the central aim of the research question, it was decided that all tests about which the coders did not agree on the categorization (a total of four) were to be excluded from the analysis.

Averages were subsequently calculated for performance on each of the Dublin descriptors. Only one test predominantly measured learning skills. This means that this measure is much less reliable than the measures for the other descriptors. Additionally, averages were calculated for performance on different types of tests (written examinations, papers, and oral examinations). A total of 83 students (out of a total of 93) had scores on all five Dublin descriptors as well as a known HSGPA. Only these students were used in the analyses.

### ***Statistical methods***

Next to descriptive statistics, a series of Pearson correlations were used to study the differences between HSGPA and the different aspects of FYGPA. To see whether these differences are significant, Steiger's *Z* (as described in Meng, Rosenthal, and Rubin 1992) was used. To study the additional effect of specific subjects in high school (native language and mathematics), next to the HSGPA, partial correlations were used.

## **Results**

### ***Descriptive statistics***

The average grades measuring the different Dublin descriptors are presented in Table 2 with their standard deviation.

The averages are quite similar across the different descriptors. The most notable difference between these statistics is the differing standard deviations. The influence of these differences in standard deviations will be, in part, negated by the standardization of scores that is a part of using Pearson correlation analyses. Still, some influence of the differences between standard deviations can remain due to the possible effect of a larger amount of tied scores when standard deviation is lower.

### ***The predictive value of high school grades***

To test the assumption that the HSGPA predicts different Dublin descriptors to different degrees, Pearson correlation analyses were conducted with the HSGPA and

Table 2. Grades for different Dublin descriptors in the first year of university.

Dublin descriptor	M	SD	95% CI	
			LL	UL
Knowledge and understanding	6.78	1.05	6.55	7.01
Applying knowledge and understanding	7.13	0.56	7.01	7.25
Making judgements	7.23	0.58	7.10	7.35
Communication	7.30	0.38	7.22	7.39
Learning skills	7.18	0.78	7.01	7.35

Note. CI, confidence interval; LL, lower limit; UL, upper limit;  $n = 83$ .

Table 3. Pearson correlations between HSGPA and performance on Dublin descriptors.

Grades	1	2	3	4	5	6
1. Knowledge and understanding	—					
2. Applying knowledge and understanding	0.52***	—				
3. Making judgements	0.55***	0.44***	—			
4. Communication	0.31**	0.20	0.26*	—		
5. Learning skills	0.48***	0.23*	0.42***	0.32**	—	
6. HSGPA	0.74***	0.35***	0.46***	0.30**	0.33**	—

Note: HSGPA, high school grade point average;  $n = 83$ .

\* $p < .05$ , two-tailed.

\*\* $p < .01$ , two-tailed.

\*\*\* $p < .001$ , two-tailed.

performance on each of the different Dublin descriptors. The results of these analyses are shown in Table 3.

There is a large difference between the amount of variability that can be predicted by the HSGPA ( $r^2$ ) for the grades that primarily measure *knowledge and understanding* and the grades that primarily measure any other Dublin descriptor. The HSGPA can predict 54.76% of the differences in grades in the *knowledge and understanding* descriptor, while it can explain only 12.25% in the *applying knowledge and understanding* descriptor, 21.16% in the *making judgements* descriptor, 9% in the *communication* descriptor, and 10.89% in the *learning skills* descriptor.

Steiger's  $Z$  analyses show that the correlation of HSGPA with performance on *knowledge and understanding* differs significantly from the correlations of HSGPA with performance on the other descriptors.

Part of the difference in predictive power of the HSGPA for different Dublin descriptors seems to overlap with the type of test (see Table 4 for the correlations between HSGPA and different types of tests), where written examinations are best predicted by HSGPA and other types to a lesser degree. These differences, certainly between papers and written examinations, are, however, much smaller (17% difference in predictive value, or even only 11% when only individual paper assignments are taken into account) than those between the different Dublin descriptors.

Table 4. Correlations between results for different types of tests and HSGPA.

Type of test	$r$	$r^2$
Written examinations	0.74***	0.55
Papers	0.62***	0.38
Individual papers	0.66***	0.44
Group assignment papers	0.30***	0.09
Oral examinations	0.33**	0.11
Individual oral examinations	0.28*	0.08
Group oral examinations	0.25*	0.06

Note:  $n = 83$ .

\* $p < .05$ , two-tailed.

\*\* $p < .01$ , two-tailed.

\*\*\* $p < .001$ , two-tailed.

Especially group assignment papers and oral examinations seem to measure markedly different things, with only a small percentage of their variability being explained by high school grades. In part, certainly for the group assignments, this is probably due to the fact that grades for individuals are mediated here by the performance of other individuals. That does not, however, explain the differences in explanatory power entirely. When group assignments are filtered out of the averages of performance on the different Dublin descriptors, the results from Table 3 hardly change. There are no group assignments as part of the *knowledge and understanding* or the *learning skills* average, and the correlation between *applying knowledge and understanding* and HSGPA only increases by 0.01 to  $r = 0.36$ ,  $n = 83$ ,  $p = 0.001$ . The correlation between HSGPA and *making judgements* even gets a little smaller, at  $r = 0.43$ ,  $n = 83$ ,  $p < 0.001$ , as does the correlation between HSGPA and the *communication* descriptor, to  $r = 0.26$ ,  $n = 82$ ,  $p = 0.018$ .

A possible explanation for the markedly lower correlation of HSGPA with oral examinations is that oral test measures less reliably. Certainly for the *communication* descriptor, this could be an issue. In the other descriptors there are no oral examinations except for *applying knowledge and understanding*, and in that case only one oral examination is part of the descriptor. This, next to the ‘exceeds the requirements’ assessment of the tests in the PAOS programme by an independent committee (QUANY and EAPAA 2011), strengthens our feeling that this possible distortion will not influence our overall conclusions, although influence on the correlation between HSGPA and the *communication* descriptor is possible.

To see whether grades for different subjects predict performance in the different Dublin descriptors, Table 5 shows the correlation between the descriptors and grades for Dutch and Mathematics A.

Table 5 shows significant correlations between grades of both subjects and grades in the *knowledge and understanding* descriptor, be it smaller correlations than with the combined grades in HSGPA. For Dutch, there are also significant correlations with *applying knowledge and understanding* and *making judgements*, and for Mathematics A with *making judgements*. A similar picture, where correlations become smaller in subsequent descriptors, is present here as well. However, when eliminating differences in HSGPA from these correlations, using partial correlation analyses, none of them remain significant. The only significant result in these partial correlations is the negative correlation between the math grade and the grades for *communication*, at  $r = -0.32$ ,  $n =$

Table 5. Correlations between high school subjects and performance on Dublin descriptors.

Dublin descriptor	Dutch		Math A	
	<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>
Knowledge and understanding	0.46***	83	0.52***	66
Applying knowledge and understanding	0.31**	83	0.16	66
Making judgements	0.29**	83	0.32**	66
Communication	0.13	83	-0.05	66
Learning skills	0.10	83	0.18	66

Note: *n* differs due to students not all having taken all subjects in high school.

\*\* $p < .01$ , two-tailed.

\*\*\* $p < .001$ , two-tailed.

63,  $p = 0.008$ , but with the lower  $n$  for students having taken this type of mathematics in high school, no regression model with HSGPA and math grade is possible that explains a significant part of the differences in performance on the *communication* descriptor. So the grades for these separate subjects cannot add significantly to the explanatory power of the HSGPA.

## Discussion

It was confirmed in this article that the HSGPA is, as has often been proven, a strong predictor for performance at university expressed as FYGPA. However, it was also shown that there are large differences between the extent to which different grades in university can be predicted by performance at high school. So the central question in this article – Are grades at the university level that measure the performance on different types of goals equally well predicted by high school GPA as a measure of previous cognitive performance? – can be answered in the negative. The performance on the Dublin descriptor of *knowledge and understanding* can be predicted more than twice as well as the performance on any of the other descriptors. Furthermore, it was shown that this difference is not primarily due to the type of tests that are part of a descriptor. Even though there are more written examinations as part of the *knowledge and understanding* descriptor, the difference between the correlations of written examinations and papers with the HSGPA is much smaller than would be expected if this would be the explanation for the differences between correlations of HSGPA with the performance on the different Dublin descriptors. This shows that the coinciding of the predictive value of the HSGPA for *knowledge and understanding* and written examinations is mostly the result of the fact that *knowledge and understanding* are more easily measured by written examinations, and next by a paper assignment, rather than that written examinations in and of themselves can better be predicted by the HSGPA. Therefore it can be concluded that the different types of assignments, including group or individual assignments, appear to be simply used to test different types of knowledge, rather than the other way around. Also, it was proven that group assignments do not dilute these results in a meaningful way, since similar results appear when they are excluded from the analysis.

The analyses involving specific individual high school subjects suggest that further study into the added predictive value of grades for specific high school subjects is not worthwhile. Even though there are correlations between the grade for native language education and mathematics on the one side and the GPA on the other, when eliminating differences in HSGPA, using partial correlation analyses, no meaningful significant correlations between the grades of these subjects and the performance on the Dublin descriptors remain. Even though the evidence suggests that *knowledge and understanding* are what is primarily graded in high school, it could be worthwhile to differentiate between types of tests at the high school level as well.

All this shows that a generic use of the FYGPA only partially delves into its possibilities of distinguishing between what university students are expected to learn and are tested on. The differentiation in grades between different levels of cognitive performance supplemented by performance in the affective domain is necessary. This is possibly also the case for performance on the psychomotor domain in programmes where such skills are appropriate, but this has not been so in the case that is reported on in this article. The Dublin descriptors, which have become increasingly important in a number of European systems of higher education, showed to be a

valuable categorization in this study. They contain different cognitive as well as affective goals that are usually part of university education (certainly in universities in the countries that are part of the Bologna process), and grades measuring performance on these descriptors can be separated into usable categories.

This also points towards the importance of an increased awareness of the different goals of university education, even the goals that are specifically part of the curriculum and its tests, in the admission process. Separate measures specifically aimed to predict performance on the different levels of the cognitive and affective domain (in some cases possibly supplemented by the psychomotor domain) would be a valuable addition to admission processes. This asks not only for acceptance of HSGPA as the best predictor for performance on the *knowledge and understanding* level of the cognitive domain, but also of acceptance that other measures will be necessary to predict other types of performance as well. No additional predictors are necessary for future performance on the lower cognitive levels, but there is more to explain.

There are limitations to this study. First, the differences in standard deviations between the Dublin descriptors (although some measures have been taken to negate this issue, its influence cannot be disproven) and the fact that the *learning skills* descriptor is only measured by one test make the analyses less robust. Still, we feel confident in concluding that the results presented in Table 3 support the conclusion that the HSGPA is a better predictor of performance in *knowledge and understanding*, than of performance in any of the other cognitive levels. The correlations between the Dublin descriptors are stronger between the different descriptors with large cognitive components, than between the cognitive and the non-cognitive descriptors. This supports the assumption that higher levels in the cognitive domain build upon each other (but also add something distinct), but that performance in the affective domain is more distinct. Second, we cannot exclude a possible influence of the fact that oral examinations measure less reliable than written tests; for the *communication* descriptor, this might have distorted the results to some extent. In the other descriptors a distortion is much less likely since no, or at most one (in the case of *applying knowledge and understanding*), oral examination measures performance on these in the study.

A third limitation is that the study was carried out in only one programme and using only Dutch HSGPA. Further research should test whether the clear differences between the predictive values of HSGPA in performance on tests that measure different types of knowledge are found in different systems of higher education and in different programmes. A more differentiated view of the FYGPA is an important nuance in many studies about selective admission and might be able to explain some of the contradictory and different findings so far (for example, the differences in the predictive power of the SAT at different universities).

All in all, this study not only supports the advice to use more than just previous grades and/or performance on standardized tests as admission instruments, but it also supports the policy to use HSGPA for selection purposes because it has again proven to be a valid predictor for (an important part of) the FYGPA. Additional instruments seem necessary to predict a large part of everything else that is expected from future students: higher order cognitive performance, as well as performance in the affective domain. Instruments developed to do just that are necessary. A more disentangled view of the FYGPA is a promising angle for finding measurable performance indicators for which predictors other than the HSGPA are necessary.

## Note

1. Higher education in the Netherlands, also at the undergraduate level, is almost entirely disciplinary, so each University has a rather large number of disciplinary 'schools' in which students choose to earn their bachelor and/or masters degree.

## References

- Anderson, L. W., and D. R. Krathwohl, eds. 2001. *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Atkinson, R. C., and S. Geiser. 2009. "Reflections on a Century of College Admissions Tests." Center for Studies in Higher Education Research & Occasional Paper Series: CSHE.4.09.
- Berlin Communiqué. 2003. *Realising the European Higher Education Area: Communiqué of the Conference of Ministers Responsible for Higher Education in Berlin on 19 September 2003*, Berlin.
- Bhattacharya, D., S. Kanaya, and M. Stevens. 2012. "Are University Admissions Academically Fair?" University of Oxford Economics Series working paper number 608.
- Bloom, B. S., ed. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay Company.
- Bridgeman, B., L. McCamley-Jenkins, and N. Ervin. 2000. *Predictions of Freshman Grade-point Average from the Revised and Recentered SAT® I: Reasoning Test*. New York: College Entrance Examination Board.
- College Board. 2014. "About the SAT." Accessed June 17. <http://sat.collegeboard.org/about-tests/sat>
- De Jong, M. 2013. "Acceptatie van selectie?. Een discoursanalyse om de beleidsverandering van loting naar selectie bij de toegang tot het hoger onderwijs te beschrijven." Master's thesis, Utrecht University.
- De Vries, A., R. E. De Vries, and M. P. Born. 2011. "Broad versus Narrow Traits: Conscientiousness and Honesty-Humility as Predictors of Academic Criteria." *European Journal of Personality* 25 (5): 336–48. doi:10.1002/per.795.
- Dunne, M., R. King, and J. Ahrens. 2013. "Applying to Higher Education: Comparisons of Independent and State Schools." *Studies in Higher Education*. doi:10.1080/03075079.2013.801433.
- Ferguson, E., D. James, and L. Madeley. 2002. "Factors Associated with Success in Medical School: Systematic Review of the Literature." *BMJ* 324 (7343): 952–57. doi:10.1136/bmj.324.7343.952.
- Friedman, B. A., and R. G. Mandel. 2009. "The Prediction of College Student Academic Performance and Retention: Application of Expectancy and Goal Setting Theories." *Journal of College Student Retention: Research, Theory & Practice* 11 (2): 227–46. doi:10.2190/CS.11.2.d.
- Gifford, D. D., J. Briceño-Perriott, and F. Mianzo. 2006. "Locus of Control: Academic Achievement and Retention in a Sample of University First-Year Students." *Journal of College Admission* 191 (Spring): 18–25.
- Hadsell, L. 2010. "Achievement Goals, Locus of Control, and Academic Success in Economics." *The American Economic Review* 100 (2): 272–76. doi:10.1257/aer.100.2.272.
- Harackiewicz, J. M., K. E. Barron, J. M. Tauer, and A. J. Elliot. 2002. "Predicting Success in College: A Longitudinal Study of Achievement Goals and Ability Measures as Predictors of Interest and Performance from Freshman Year through Graduation." *Journal of Educational Psychology* 94 (3): 562–75. doi:10.1037/0022-0663.94.3.562.
- Jacobs, M. 2010. "A Framework for the Placement of University Students in Science Programmes." PhD diss., University of the Free State.
- Joint Quality Initiative Group. 2004. *Shared 'Dublin' Descriptors for Short Cycle, First Cycle, Second Cycle and Third Cycle Awards: A Report from a Joint Quality Initiative informal Group*. [http://www.nvao.net/page/downloads/Dublin\\_Descriptors.pdf](http://www.nvao.net/page/downloads/Dublin_Descriptors.pdf)
- Kehm, B. M. 2010. "Quality in European Higher Education: The Influence of the Bologna Process." *Change: The Magazine of Higher Learning* 42 (3): 40–46. doi:10.1080/00091381003704677.



- Komarraju, M., S. J. Karau, and R. R. Schmeck. 2009. "Role of the Big Five Personality Traits in Predicting College Students' Academic Motivation and Achievement." *Learning and Individual Differences* 19 (1): 47–52. doi:10.1016/j.lindif.2008.07.001.
- Krathwohl, D. R., B. S. Bloom, and B. B. Masia. 1964. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Company.
- Kuncel, N. R., and S. A. Hezlett. 2010. "Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions." *Current Directions in Psychological Science* 19 (6): 339–45. doi:10.1177/0963721410389459.
- Kwaliteit in Verscheidenheid Act of 2013, Dutch Bulletin of Acts, Orders and Decrees 2013. nr. 298 and 311.
- McKenzie, K., K. Gow, and R. Schweitzer. 2004. "Exploring First-Year Academic Achievement through Structural Equation Modelling." *Higher Education Research and Development* 23 (1): 95–112. doi:10.1080/0729436032000168513.
- Meng, X. L., R. Rosenthal, and D. B. Rubin. 1992. "Comparing Correlated Correlation Coefficients." *Psychological Bulletin* 111 (1): 172–75. doi:10.1037/0033-2909.111.1.172.
- Petska, K. S. 2006. "Using Personality Variables to Predict Academic Success in Personalized System of Instruction." PhD diss., The University of Nebraska-Lincoln.
- QANU (Quality Assurance Netherlands Universities). 2012. *Guideline for Writing a Self-Assessment Report for a Limited Programme Assessment*. [http://www.qanu.nl/sites/default/files/bestanden/files/Guideline\\_writing\\_SER\\_limited\\_assessment\\_version\\_103\\_0.pdf](http://www.qanu.nl/sites/default/files/bestanden/files/Guideline_writing_SER_limited_assessment_version_103_0.pdf)
- QANU (Quality Assurance Netherlands Universities), and EAPAA (European Association for Public Administration Accreditation). 2011. *Assessment of the Degree Programmes in Public Administration and Organization Science at Utrecht University*. [http://search.nvao.net/files/517a7369da22a\\_rapport%20UU%20wo%20Bestuurs-%20en%20Organisatiewetenschap.pdf](http://search.nvao.net/files/517a7369da22a_rapport%20UU%20wo%20Bestuurs-%20en%20Organisatiewetenschap.pdf)
- Reumer, C., and M. Van der Wende. 2010. "Excellence and Diversity: The Emergence of Selective Admission Policies in Dutch Higher Education – a Case Study on Amsterdam University College." Center for Studies in Higher Education Research & Occasional Paper Series: CSHE.15.10.
- Richardson, M., C. Abraham, and R. Bond. 2012. "Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis." *Psychological Bulletin* 138 (2): 353–87. doi:10.1037/a0026838.
- Robbins, S. B., K. Lauver, H. Le, D. Davis, R. Langley, and A. Carlstrom. 2004. "Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis." *Psychological Bulletin* 130 (2): 261–88. doi:10.1037/0033-2909.130.2.261.
- Sackett, P. R., M. J. Borneman, and B. S. Connelly. 2008. "High-Stakes Testing in Higher Education and Employment: Appraising the Evidence for Validity and Fairness." *American Psychologist* 63 (4): 215–27. doi:10.1037/0003-066X.63.4.215.
- Schmitt, N. 2012. "Development of Rationale and Measures of Noncognitive College Student Potential." *Educational Psychologist* 47 (1): 18–29. doi:10.1080/00461520.2011.610680.
- Shultz, M. M., and S. Zedeck. 2012. "Admission to Law School: New Measures." *Educational Psychologist* 47 (1): 51–65. doi:10.1080/00461520.2011.610679.
- Siu, E., and H. I. Reiter. 2009. "Overview: What's Worked and what hasn't as a Guide towards Predictive Admissions Tool Development." *Advances in Health Science Education* 14 (5): 759–75. doi:10.1007/s10459-009-9160-8.
- Soares, J. A. 2012. "The Future of College Admissions: Discussion." *Educational Psychologist* 47 (1): 66–70. doi:10.1080/00461520.2011.638902.
- Stemler, S. E. 2012. "What should University Admissions Tests Predict?" *Educational Psychologist* 47 (1): 5–17. doi:10.1080/00461520.2011.611444.
- Te Wierik, M. L. J., J. Bieshuizen, and W. Van Os. 2014. "Career Guidance and Student Success in Dutch Higher Vocational Education." *Studies in Higher Education*. doi:10.1080/03075079.2014.914905.
- Uppal, N., and S. K. Mishra. 2013. "Moderation Effects of Personality and Organizational Support on the Relationship between Prior Job Experience and Academic Performance of Management Students." *Studies in Higher Education*. doi:10.1080/03075079.2013.777411.
- Urlings-Strop, L. C., T. Stijnen, A. P. Themmen, and T. A. Splinter. 2009. "Selection of Medical Students: A Controlled Experiment." *Medical Education* 43 (2): 175–83. doi:10.1111/j.1365-2923.2008.03267.x.

- Wintre, M. G., B. Dilouya, S. M. Pancer, M. W. Pratt, S. Birnie-Lefcovitch, J. Polivy, and G. Adams. 2011. "Academic Achievement in First-Year University: Who Maintains their High School Average?" *Higher Education* 62 (4): 467–81. doi:10.1007/s10734-010-9399-2.
- Zimdars, A. 2010. "Fairness and Undergraduate Admission: A Qualitative Exploration of Admissions Choices at the University of Oxford." *Oxford Review of Education* 36 (3): 307–23. doi:10.1080/03054981003732286.