University Teacher Judgments in Problem-Based Learning in Higher Education: Their Accuracy and Reasoning

Lisette Wijnia[a,b*], Sofie M. M. Loyens[a,b], Eva Derous[c], & Henk G. Schmidt[a]

[a]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, the Netherlands

[b] Roosevelt Center for Excellence in Education, University College Roosevelt, Utrecht University, Lange Noordstraat 1, 4330 AB Middelburg, the Netherlands

[c]Department of Personnel Management, Work & Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium

*Correspondence concerning this article should be addressed to Lisette Wijnia, Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, the Netherlands. E-mail: lisettewijnia@gmail.com

University Teacher Judgments in Problem-Based Learning in Higher Education: Their Accuracy and Reasoning

Abstract

This study investigated the accuracy of 14 university teacher's judgments. Early in the first year, university teachers rated the chance each university student in their group would successfully complete their first year as well as the entire bachelor's program. Results show that university teachers' chance ratings were predictive of actual academic success. However, they were more accurate in predicting success than failure. Moreover, results revealed that university teachers mostly built upon their observations of university students' engagement and motivation, instead of students' cognitive ability in their judgments. Unsuccessful university students received relatively more negatively framed statements than successful students did.

Keywords: Teacher judgments; judgment accuracy; problem-based learning; higher education.

1. Introduction

Preventing dropout and study delays are major concerns in higher education. According to the Organization for Economic Co-Operation and Development (OECD, 2013), around 30% of students who enter a higher education program leave without a degree. Dropout rates in the Netherlands are similar to the OECD average (OECD, 2013; Educational Inspectorate, 2013). However, not only dropout is a concern, many students experience study delays longer than one year. In the Netherlands, only 26.40% of university students obtain their bachelor's degree on time (i.e., within 3 years), and 48.60% of university students graduate with a one-year delay (Educational Inspectorate, 2013). Study delays and dropout can be both time-consuming and costly for students, as well as for institutes of higher education. For example, in several European countries, the amount of funding universities receive from the government depends on the number of students who graduate (Author, 2014a; Hovdhaugen, 2009).

A majority of university students who leave higher education without any degree do so during or immediately after their first bachelor's year (Tinto, 1993, 1998). It is therefore important to target dropout at an early stage of university students' academic careers. In this study, we aim to investigate whether university teachers can identify first-year university students at risk of dropout or delays during the bachelor's program and which student characteristics university teachers perceive as important for academic success.

2. The Accuracy of Teacher Judgments

Teachers judge students' performance or behavior (Südkamp, Kaiser, & Möller, 2012) for diverse purposes, such as formal assessment, referral decisions (e.g., remedial teaching or acceleration), and instructional decisions (e.g., selection of tasks, determining difficulty levels, and organization of learning; Abidin & Robinson, 2002; Begeny, Eckert, Montarello, & Storie,

2008; Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006; Gerber & Semmel, 1984; Südkamp et al., 2012). It therefore is important to judge students in an accurate way.

Judgment accuracy is typically defined as the correlation between teacher judgments of students' academic achievement and students' actual academic achievement, such as on standardized tests (e.g., Südkamp et al, 2012). Jussim (1989, 1991) further argued that only when teacher judgments predict student behavior or achievement, without causing it, one can speak of accuracy. Accurate judgments are based on relevant background information, such as students' ability. However, according to Jussim's (1991) *reflection-construction model* teacher judgments could be inaccurate when instigated by expectancy effects and perceptual bias. An example of teacher expectancy effects are self-fulfilling prophecies, where teacher judgments about students will change student behavior so that the initial teacher expectation is confirmed. Perceptual bias takes place when teachers interpret students' performance in ways that are consistent with their initial beliefs or ideas about a student's capabilities and competencies regardless of any independent, objective assessment of students' capabilities and competencies (like with standardized tests). Such perceptual biases can be instigated by stereotypical beliefs (e.g., social economic status, ethnicity, culture, gender) teachers believe are associated with students' study performance.

Studies that examined the magnitude of judgment accuracy, teacher expectation effects, and perceptual bias, concluded that self-fulfilling prophecy effects are often small and that teachers are mostly accurate in judging student performance (Jussim & Harber, 2005; Trouilloud, Sarrazin, Martinek, & Guillet, 2002). Moreover, two meta-analyses demonstrated that teacher judgments of students' performances are quite accurate (Hoge & Coladarci, 1989; Südkamp et al., 2012). For example, the meta-analysis by Südkamp et al. (2012) of research conducted from

1989 to 2009 resulted in a mean correlation of .63 (range r = -.03 to r = .84) across 75 studies. Whereas studies in these meta-analyses predominately involved samples from kindergarten or elementary school, only a minority also considered secondary education. However, even less is known about the accuracy of teacher judgments in higher education settings. This is remarkable, given the high dropout rates in higher education and associated costs (OECD, 2013). Moreover, fundamental judgments university teachers make about the quality of student work are often subjective decisions, such as judging the extent to which an objective is met (Sadler, 2005). One reason for this knowledge gap, perhaps, might be the larger class sizes, making it more difficult to investigate teacher judgment accuracy. Therefore, gaining more insight into the accuracy of teacher judgments in higher education is important.

Interestingly, the few studies that considered judgements in higher education reported mixed results. For example, Chamorro-Premuzic and Furnham (2003) demonstrated in one study that university teachers' exam predictions were related to university students' actual exam scores (explaining 6% of the variance), whereas in a second study no significant associations were found. Kaufman and Hansell (1997) and Whitfield and Xie (2002) found positive associations between university teacher ratings' of knowledge and students' actual exam scores. Nevertheless, teacher ratings could only explain little variance in actual exam scores (Kaufman & Hansell, 1997) and university teachers were likely to overestimate students' knowledge base, especially for students situated in the bottom 25% of the class (Whitfield & Xie, 2002).

Other studies were more positive about the accuracy of teacher judgments. Van de Watering and Claessens (2003) demonstrated that university teachers' classification of their first-year law students as barely, moderately, or highly competent corroborated with actual exam scores. Finally, Author (2014b) found that teacher judgments made early in the first bachelor's

year were predictive of university students' academic success in that first bachelor's year as well as across the entire bachelor's program (explaining 10-22% of the variance). However, effects were not unequivocal: Results indicated that university teachers were better in predicting academic success than failure, which warrants further investigation. Interestingly, this asymmetrical effect is in line with studies conducted in primary and secondary education that indicated that primary and secondary teachers' judgment accuracy was higher for high-achieving students (e.g., Demaray & Elliott, 1998) and that teachers were better at predicting who would not develop learning difficulties than those who would develop learning difficulties (Flynn & Rahbar, 1998; Gijsel, Bosman, & Verhoeven, 2006; Taylor, Anselmo, Foreman, Schatschneider, & Angelopoulos, 2000).

3. How Do Teachers Make Their Judgments?

To gain more insight into teacher judgment accuracy, it is important to learn more about the type of information teachers use when making judgments about students' performance and ability. Teachers may judge students' achievements on the basis of student characteristics other than ability, which may affect the overall accuracy of their judgments. For example, previous research revealed that 6th grade teachers' and university teachers' judgments of achievement were influenced by students' behavioral engagement in class (e.g., the number of questions a student asks in class, absenteeism in class; Chamorro-Premuzic & Furnham, 2003; Kaiser, Retelsdorf, Südkamp, & Möller, 2013). This was even the case in two experimental studies by Kaiser et al. (2013). In these experiments, teachers candidates participated in a computer simulation of a classroom situation in which they interacted with virtual elementary and secondary school students. These students had experimentally manipulated levels of achievement and engagement in terms of the proportion of correct answers and participation in class.

Afterward, teacher candidates were asked to judge students' achievement and engagement. Although, in these simulations, the correlation between students' actual engagement and actual achievement was constrained to zero, results demonstrated that teachers inaccurately based part of their judgments of students' achievement levels on students' displayed behavioral engagement in class. Kaiser et al. (2013) suggested that teachers might have taken the collinearity of engagement and actual achievement into account when making their judgments. That is, teachers might assume that high engagement and high achievement go hand in hand.

If teacher judgments are indeed influenced by students' engagement because of its assumed relationship with achievement, it is possible that other non-intellectual factors play a role as well. Although little is known about how teachers' perceptions of student characteristics influence teacher judgments' of students' achievement, there are some indications that teachers take perceptions of individual differences, such as assumed personality and motivation, into account when predicting students' final grades. For example, Doherty and Conolly (1985) demonstrated that primary school teacher judgments were influenced by pupils' perceived tidiness. Further, Urhahne (2015) found that secondary school students who were underestimated by their teachers were perceived as less motivated than overestimated students were.

4. Present Study and Hypotheses

The Dutch higher education system consists of two types of institutes: research universities and higher vocational education offered at universities of applied sciences. Both systems have a bachelor's-master's degree structure. The current study was conducted at a research university. In this study, we investigated the predictive value of 14 university teacher's judgments during the first course of a problem-based, bachelor's psychology program (3 full-time years) for university students' (N = 250) successful completion of their first year as well as

the entire bachelor's program. Problem-based learning (PBL) is characterized by student-centered, collaborative learning in small groups of 10-11 university students under the guidance of a university teacher (Barrows, 1996). Realistic problems are used as the starting point. After students are presented with a problem description, the PBL-cycle generally consists of three phases: (1) the initial discussion phase, (2) the self-study phase, and (3) the reporting phase. During the initial discussion phase students discuss the problem description and come up with possible explanations for the problem. In doing so, they use their prior knowledge and common sense. Because prior knowledge is typically insufficient to understand the problem situation, students formulate research questions for further self-study. During the self-study phase, students consult scientific literature to come up with answers to their research questions. In the report phase, students meet again (same group constellation) to discuss their literature findings to come to an answer to the learning issues. PBL was first developed in medical education in Canada, but is now applied in many different disciplines and institutes around the world (Author, 2009).

In PBL, the university teacher has several roles, such as guiding and facilitating the learning process in group meetings, monitoring the group process, and providing feedback to each student's individual learning process (Author, 2012b). These teacher roles and the interactive nature of PBL might enable them to provide useful information concerning university students at risk of study delays and dropout. Accurate judgments can only be made to the extent that relevant behavioral information is available to and detected by a judge (Funder, 2012).

The study consists of both a two parts that will be discussed separately for the sake of clarity. The first part focuses on the accuracy of university teachers' judgments made during the first course of the academic year. As yet only a few studies examined the accuracy of teacher judgments in higher education, the first part allowed to examine whether the asymmetrical

results in accurate judgment of academic success versus failure could be replicated. Specifically, university teachers were asked to judge the chance (0-100%) that students would successfully complete the first bachelor's year, the bachelor's degree, and whether students would obtain their degree in time. In addition, university teachers were asked to judge students' observed learning activities in class, such as level of preparation, participation, knowledge base, self-confidence, and interest. With respect to this part of the study, we expect that university teachers' predictions of academic success will be positively correlated with their judgments of engagement in learning activities in class *(Hypothesis 1*). Moreover, we hypothesize that teacher chance ratings are predictive of students' actual academic success during the first year and the entire bachelor's program *(Hypothesis 2).* However, we hypothesize that university teachers are better in identifying successful university students than unsuccessful university students *(Hypothesis 3).*

In the second part of the study we examined university teacher's reasoning behind their judgments using a qualitative approach. University teachers were asked to motivate their chance ratings by describing on what information or observation during the group meetings they had based their judgments. Until today it is unclear which information teachers actually use to base their judgments on. By specifically investigating the type of information that teacher judgments are based on, we address the recent call from Kaiser et al. (2013) to identify student characteristics other than engagement that might moderate or even bias judgment accuracy. We expect that, in addition to variables related to students' cognitive ability (such as intelligence or prior educational attainments), characteristics related to demographic variables and non-intellectual constructs such as personality and motivation will also be mentioned *(Hypothesis 4).* Finally, we will explore whether university teachers' observations differ for university students who successfully finished the first year or failed.

5. Part 1: Accuracy of University Teachers' Judgments

5.1 Method

5.1.1 Participants and design. Participants were university teachers and university students from a full-time 3-year bachelor of psychology program of a research university (20,941 students in total) in a major city in the Netherlands. The psychology program was founded in 2001. In the bachelor's program under study, the first year consists of 8 successive 5-week courses. Each 5-week course consists of 9 3-h face-to-face discussion group meetings, 5 optional 2-h lectures, and usually 4 mandatory 3-h practical meetings. Each week 30 hours are reserved for (individual) self-study. Each course deals with a different discipline of psychology (e.g., social psychology, personality psychology, clinical psychology). For each 5-week course, students are randomly assigned to a new discussion group consisting of 10-11 students and a teacher. Due to the active learning philosophy of the psychology program maximally 290 students can enroll in the first year each year.

After the third week of the first course (i.e., social psychology), all university teachers were asked to fill out a questionnaire for all the first-year university students in their discussion groups. In terms of ethical approval, permission for this study was given by the educational director and the course coordinator of the course in which the questionnaire took place. With permission of the course coordinator of the first course, university teachers were asked if they were willing to participate in a study investigating whether teacher judgments made early on in the first year were predictive of academic success. University teachers' participation was voluntary and university teachers received a 10 euro gift certificate for participation. Please note that none of researchers were teaching in the first year of the bachelor´s program or involved in any other way.

After obtaining their informed consent, university teachers received a questionnaire to fill out. In this questionnaire, university teachers had to indicate each student's chances of success (0-100%) and teachers were asked to motivate their predictions. Finally, university teachers were asked to rate university students' engagement in class. Fourteen (1 male, 13 female) of the 15 first-year university teachers participated in the study. University teachers were on average 29.36 (*SD* = 8.17) years old. All university teachers had a master's degree in psychology and had received a three-day training in guiding PBL discussion groups. We asked each teacher to fill out their amount of experience in guiding PBL discussion group. Four university teachers were novices, as they had no teaching experience in facilitating a PBL discussion groups prior to this course. All other university teachers had teaching experience in PBL, however, three of them had less than 1 year of experience, 1 university teacher had between 1-2 years of experience, and 5 university teachers had more than two years of experience in guiding PBL discussion meetings. One university teacher failed to indicate her amount of experience in guiding PBL discussion meetings. 14.29% (*n* = 2) of the university teachers had a full-time contract as an academic teacher, whereas the other teachers (85.71%, *n* = 12) had a part-time contract.

The 14 university teachers made predictions about 250 first-year university students. University students were on average 19.69 years old (*SD* = 2.09; 75.60% female). Information about university students' study progress was obtained through the university's administrative office with permission of the educational director of the bachelor's program.

5.1.2 Teacher chance ratings. University teachers were asked to predict the chance each student would pass the first year, complete the entire bachelor's program, and would obtain their bachelor's degree in time (i.e., within three years). University teachers had to indicate these chances of success on a scale from 0-100%, by marking a university student's chance of success

on a bar of 10 centimeters long. For example, when the bar was marked at 5 centimeters, this indicated a 50% chance of success.

5.1.3 Teacher ratings of engagement. University teachers were asked to rate first-year bachelor's students' engagement in class on a scale from 0 to 100%. The scale consisted of six items: preparation, active participation during brainstorm/problem analysis phase, active participation during the reporting phase, displayed understanding of the subject matter in group meetings, displayed self-confidence, and observed interest in the subject matter. The items were, in part, based on the scale for observed learning activities as described by Author (2007), and on literature on engagement during learning activities (Fredricks, Blumenfeld, & Paris, 2004; Reeve, 2012). The items regarding preparation and participation are indicators of behavioral engagement. Observed interest and self-confidence refer to emotional engagement, whereas the level of understanding can be seen as an indicator of cognitive engagement (see Reeve, 2012). Principal component analysis demonstrated that the six items loaded on one factor "teacher judgments of engagement", explaining 73.41% of the variance with factor loadings ranging between .80 and .94. In addition, the scale had a satisfactory reliability (Cronbach's $\alpha = .93$).

5.1.4 Successful completion of the first year. In the first year of the curriculum under study, university students need to obtain 60 European Credits (EC). Forty of the 60 EC represent a "knowledge" cluster. First-year bachelor's students obtain these 40 EC if their average grade on 8 course tests is a 6.0 or higher (on a scale from 0 to 10). In addition, none of the course tests should have received a grade lower than 4.0. The remaining 20 EC can be obtained through practical assignments associated with each course (e.g., academic writing skills, research skills, clinical communication skills, presentation skills), representing a "skills" cluster. Again, university students receive these 20 EC when the average grade associated with the practical

assignments is a 6.0 or higher and none of the grades is lower than a 4.0. Bachelor's students have successfully completed the first year if they have earned all 60 EC (coded as 1). They fail if they have obtained less than 60 EC (coded as 0).

5.1.5 Successful completion of the bachelor's program. To complete the bachelor's degree, university students need to obtain 180 EC (60 EC each year, three years in total). University students' academic success in the bachelor program was coded as "0" if they had left the program without a degree, "1 "when they experienced a delay (i.e., were still enrolled but earned less than the 180 EC after three years), or "2" when they had obtained their bachelor's degree.

5.2 Results and Discussion

5.2.1 Hypothesis 1: Association between teacher judgments. On average, university students received high chance ratings of success and high ratings of observed engagement during group meetings, ranging between 67.52% (*SD* = 21.81) for obtaining the bachelor´s degree in time to 73.04% (*SD* = 19.16) for successfully completing the first year. In support of Hypothesis 1, there were high correlations between university teachers' judgments of engagement and their ratings of successfulness in the first year ($r = .83$, $p < .001$), bachelor´s program ($r = .82$, $p < .001$), and obtaining the bachelor´s program in time ($r = .73$, $p < .001$), suggesting that these judgments are highly intertwined in their eyes. Due to potential multicollinearity, in subsequent analyses we only examined the predictive value of university teachers' chance ratings.

5.2.2 Hypothesis 2: Judgment accuracy. To investigate Hypothesis 2, we examined whether university teacher chance ratings were predictive of actual success in the first-year and completion of the bachelor's degree. Given the relatively small sample size, we did not conduct multilevel analysis but corrected the standard errors for cluster level by conducting logistic

regressions in Mplus 7.31 with the robust maximum likelihood estimator, type is complex (see Hox, Maas, & Brinkhuis, 2010, Muthén & Muthén, 2012).

First, university students' first year success was investigated. Actual results indicated that 58 bachelor's students (23.20%) failed to pass the first year and 192 (76.80%) students obtained the mandatory 60 EC and successfully completed the year. The intraclass correlation indicated about 5% of the variability in successful completion of the first year could be described as between group variability. As can be seen in Table 1, university teacher ratings significantly predicted successful completion of the first year explaining 13% of the variance in successful completion. The corresponding odds ratio was 1.04. An odds ratio (OR) larger than 1 indicates that when teacher ratings increases, the chance of passing the first year increases as well, supporting Hypothesis 2.

With respect to bachelor' program success, 106 university students (42.40%) obtained their bachelor's degree in time, 79 students (31.60%) were still enrolled but experienced some study delay, and 65 students (26%) left the program without a degree. We investigated the predictive value of university teachers' chance ratings of success and completion of the bachelor's program (graduated = 1, otherwise = 0). About 2% of the variability in successful completion of the bachelor's degree in time could be described as between-teacher group variability. In support of Hypothesis 2, university teachers' chance ratings of obtaining the bachelor's degree in time made early in the first year significantly predicted whether students obtained their degree or not, explaining 5% of the variance. The corresponding odds ratio was 1.02 indicating that when teachers change ratings increased, the chance of obtaining the degree within three years increased as well.

Moreover, we conducted a multinomial logistic regression to investigate whether teacher chance ratings of completing the bachelor's program could differentiate between students who left the program without a degree or experienced a study delay versus students who obtained their bachelor's degree within three years (Table 2). In support of Hypothesis 2, students who left the program without a degree received lower chance ratings than students who completed the program in time (OR = 0.96). In addition, students who were enrolled but needed more than three years to complete the program received lower chance ratings than students who completed the program in time (OR = 0.98).

5.2.3 Hypothesis 3: University teachers are better able to predict success. To examine Hypothesis 3, students were first classified as *"misses*," *"hits*," *"correct rejections*," and *"false alarms"* (Table 3). Hits and correct rejections are correctly classified university students (Gijsel et al., 2006). A *hit* concerns a student who was predicted to have a high chance of success and indeed passed the first year or bachelor's degree, whereas a *correct rejection* concerns a student who was predicted to have a low chance of success and eventually failed the first year or bachelor's program. A *miss* refers to students who were predicted to have a low chance of success, but did pass the first year or obtain a bachelor's degree, whereas a *false alarm* refers to students who failed/dropped out, but were predicted to have a high chance of success. Based on these classifications the specificity and sensitivity of university teacher judgments could be calculated. *Specificity* refers to the proportion of bachelor's students who were correctly identified by teachers to fail (i.e., correct rejection) given all those who failed. *Sensitivity* concerns the proportion of bachelor's students who were correctly identified to pass the first year or to complete the bachelor degree (i.e., hit) relative to all students who successfully completed the first year or bachelor's degree.

With respect to first-year success, analyses including all bachelor's students who received chance ratings above and below 50% for the first year resulted in a specificity of .24 and a sensitivity of .93. Subsequently, we decided to recalculate the specificity and sensitivity of university teachers' judgments when only the bachelor students with more extreme teacher ratings were included: i.e., the university students who received chance ratings of 75% or higher to complete the first-year and the university students who received a chance rating of 25% or lower. Sensitivity was high (.99). In total only 6 university students received a chance rating below 25%. Although most of these students indeed failed the program, specificity was low (.19) because most of university students who failed the first year were not correctly classified.

For bachelor´s program success, we calculated how well teachers could discriminate between the two most extreme categories: successful university students who obtained their degree without study delays and those students who left the program without a degree. Analyses including all bachelor's students who received a chance rating above and below 50% to complete the bachelor's program success resulted in a specificity of .23 and a sensitivity of .92. When only the bachelor's students with more extreme teacher ratings were included (i.e., 75% or higher and 25% or lower), we obtained a perfect score for sensitivity. However, specificity was again low (.19). Therefore, in support of Hypothesis 3, university teachers were better at correctly identifying successful rather than unsuccessful bachelor's students.

6. Part 2: Reasoning of Teacher Judgments

6.1 Method

6.1.1 Participants and design. University teachers (N = 14) were asked to motivate their chance ratings by indicating for each student which observations, factors or other information sources they used to base their judgments upon**.** Most university teachers only provided an

elaborate explanation for the first-year chance rating. Therefore, only responses regarding first-year success were coded and analyzed. The question was answered for 238 of the 250 university students in our sample (95.20%), who were on average 19.70 years old (*SD* = 2.13) and of which 74.79% were female. Fifty-six university students did not successfully complete their bachelor's year, whereas 182 university students successfully obtained the required 60 EC.

6.1.2 Coding scheme for teacher-provided reasons for success. To analyze university teachers' reasoning behind their judgments, a coding scheme was developed. Based on Richardson, Abraham, and Bond (2012) our coding scheme consisted of three broad categories: (A) demographic characteristics, (B) intellectual factors, and (C) non-intellectual factors. In addition, for each statement it was coded whether its influence on academic success was perceived as being a negative, neutral, or positive with respect to first-year success (see Appendix A for an overview of the coding scheme).

Statements reflecting university students' age, gender, or ethnicity were included in the demographic category. We included this category, because prior research has demonstrated that gender, age, ethnicity, and socioeconomic background has been associated with academic achievement in higher education (e.g., Richardson et al., 2012; Van den Berg & Hofman, 2005), but might also induce bias (cf. Jussim & Harber, 2005).

The category intellectual factors consisted of three subcategories: (B.1) general cognitive ability, (B.2) verbal ability, and (B.3) prior educational experiences and attainments. The subcategory general cognitive ability included statements concerning students' cognitive capabilities (e.g., smart, intelligent) and understanding of the subject matter, whereas the verbal ability category reflected specific statements about the quality of students' spoken and written language skills. Prior research demonstrated that university students' verbal language skills were

associated with academic achievement in a PBL, psychology, bachelor's program (Author, 2012a). It is likely that university teachers report on students' verbal abilities, because these skills are important and can be observed in the interactive PBL environment under study. Finally, university teachers' statements about students' prior educational attainments and experiences were included as a subcategory, as research has indicated that prior educational attainment are determinative for educational success in higher education and are often used as a proxy for intelligence (e.g., Central Bureau for Statistics, 2009; Jansen, 2004; Jansen & Bruinsma, 2005; Richardson et al., 2012).

Finally, the non-intellectual factors' category was divided in nine subcategories. Research has indicated that variables such as motivation, personality, and engagement in class are important predictors of academic achievement in higher education (e.g., Author, 2012a; Poropat, 2009; Richardson et al., 2012). Three subcategories were devoted to statements referring to stable personality traits: (C.1) personality – not specified, (C.2) extraversion-introversion, and (C.3) conscientiousness. Extraversion-introversion was included because of the collaborative nature of the PBL environment in which university teachers observe interactions between students and might therefore be inclined to take traits such as talk activeness, socializing, and initiative taking into account. In addition, extraversion was associated with teacher predictions of exam grades in previous research (Furnham & Medhurst, 1995). Conscientiousness (e.g., organization, planning, and discipline) was included due to its consistent relationship with academic achievement (Author, 2012a; Poropat, 2009). Other subcategories included: (C.4) collaboration readiness, (C.5) (in)stability, (C.6) study priority, (C.7) motivation and interest, and (C.8) preparation and participation in group discussions. Collaboration readiness was included because of the collaborative nature of PBL. It included statements concerning help-seeking

behavior, listening skills, and collaboration with other students. The subcategory (in)stability reflects both statements concerning general nervousness and statements concerning test/presentation anxiety. Statements concerning balancing work, study-related activities, and social life were included in the subcategory study priority. The subcategory motivation/interest reflects statements about reasons for studying and expressed interest. The subcategory preparation and participation to group discussions includes statements such as the level of preparation for group meetings but also participation such as one's contributions and questions during the meetings (see Author, 2007, 2012a). Finally, a rest category was included: (C.9) other non-intellectual factors and consisted of, for example, personal or psychological problems of students.

University teachers' answers were divided in meaningful and distinguishable units. For example, one university teacher motivated her prediction by stating: "Very motivated and enthusiastic, always prepared, active during discussions, smart, reads more than one resource." This answer was subdivided and coded as follows: very motivated and enthusiastic (C.7, positive); always prepared (C.8, positive); active during discussion (C.8, positive); smart (B.1, positive); reads more than one resource (C.8; positive). Two independent raters, resulting in a kappa coefficient of .84, coded fifty percent of the teachers' answers. Differences in scoring were resolved through discussion.

6.2 Results and Discussion

6.2.1 Hypothesis 4: University teachers' reasoning. Coding of teacher-provided reasons resulted in 876 separate units (i.e., identified reasons for success). The units mostly reflected positive student characteristics (68.15%) when compared to negative (30.25%), or neutral characteristics (1.60%). Table 4 presents an overview of the frequency in which different

categories were mentioned. Hypothesis 4 stated that characteristics related to demographic variables and non-intellectual constructs such as personality and motivation will also be mentioned in addition to university students' cognitive ability. In support of Hypothesis 4, only 192 (21.92%) of the 876 identified units were related to intellectual factors. Non-intellectual characteristics were more often mentioned (77.40%), whereas demographic characteristics were hardly mentioned (< 1%). A chi-square test revealed that statements more often reflected non-intellectual factors than intellectual factors, $\chi^2(1) = 271.49$, $p < .001$.

With respect to "demographic characteristics," only statements about university students' age were mentioned as a possible influencing factor of first-year success (see Appendix A for examples). The majority of statements within the category "intellectual factors" concerned statements about general cognitive ability (79.17 %), such as a student's ability to understand the subject matter, whereas statements about verbal abilities (8.33%) and prior educational experiences were mentioned less often (12.50%).

Almost half of the non-intellectual statements (46.76%) concerned the quantity and quality of university students' preparation for and participation to group discussions. In fact, observed preparation and participation reflected more than a third of all teacher-provided reasons. Other non-intellectual factors that were often mentioned were motivation and interest (23.30% of non-intellectual statements and 18.04% of all statements), and personality characteristics such as extraversion-introversion (9.88% of non-intellectual statements and 7.65% of all statements) and conscientiousness (7.52% of non-intellectual statements and 5.82% of all statements). Overall, these results suggest that university teachers believe that indicators of general cognitive ability, observed level of preparation and participation, expressed motivation

and interest, and personality factors such as extraversion and conscientiousness are important student characteristics that influence a university student's chance to complete the first year.

6.2.2 Differences between successful and unsuccessful bachelor's students. Finally, we examined the positive and negative statements in more detail for the bachelor's students who successfully completed the first year and for the students who failed the first-year. As can be seen in Table 5, the university students who failed their first bachelor's year received relatively more negatively framed statements than the students who did successfully complete the first year ($z = 14.21$, $p < .001$). Most negatively stated teacher-provided reasons for the students who failed the first year were categorized as non-intellectual factors, such as statements concerning the quality and quantity of the preparation for, and participation to, group meetings.

7. General Discussion

According to the reflection-construction model, teacher judgments are accurate if they predict students' actual performance, without causing it and when the judgments are based on relevant background information, such as students' ability (Jussim, 1991). Previous research among primary and secondary school teachers has demonstrated that teacher judgments are relative accurate. However, little is known about the accuracy of university teachers' judgments. In the present study, we investigated whether university teachers' judgments of bachelor's students' chance to successfully complete the first academic year were accurate. In addition, we investigated whether university teachers' chance predictions were influenced by teachers' perceptions of student characteristics. Indeed, Kaiser et al. (2013) call for more research on accuracy of teacher judgments as they suggested that teacher judgments of performance may be influenced by students' actual performance, but also by student characteristics such as engagement in class (Kaiser et al., 2013). By asking university teachers to indicate their

reasoning behind their judgments, we addressed the research call from Kaiser et al. (2013) to further unearth which perceptions of student characteristics might moderate or bias teacher judgment accuracy.

7.1 Are Teacher Judgments Accurate?

In line with prior research and Hypothesis 2, we found that university teachers' chance ratings made early in the first year were predictive of actual academic success during the first year and entire bachelor's program, explaining 5-13% of the variance (Author, 2014b; Hoge & Coladarci, 1989; Südkamp et al., 2012). In addition, sensitivity (range .93 – 1.00) and specificity (range .19 - .24) analyses indicated that university teachers were better at identifying successful university students relative to unsuccessful students. This is consistent with earlier findings (e.g., Author, 2014b) and calls into question how accurate university teachers really are in predicting students' academic failure. First, the base rate of students' success is higher than that of failure, which might have influenced the results: The majority of the university students in our sample (76.80%) successfully completed their first year and 42.40% of the students obtained their bachelor's degree in time. Second, university teachers might be inclined to give students the benefit of the doubt when judging their performance and are less inclined to give low predictions (Whitfield & Xie, 2002). In support of this, we found that when university teachers had to indicate on which information or observations teachers based their judgments, they mostly reported characteristics that were positive indicators of success. Urhahne (2015) demonstrated that although 6th grade teachers tend to overestimate student achievement, the class-wise rank component between teacher performance expectations on a language test and actual student achievement was high. Therefore, teachers might be better at predicting the relative, rather than the absolute level of students' performance. In support, our study demonstrated university

teachers' chance ratings for students who successfully completed the bachelor's degree in time were significantly higher than for students who experienced a study delay or dropped out.

7.2 How do Teachers Make Their Judgments?

In our study, we demonstrated that university teachers' judgments are often influenced by or associated with their perception of non-cognitive student characteristics. We revealed that the correlation between university teachers' chance ratings and their ratings of engagement during group meetings was high (r = .73 - .83). This suggests that teachers are potentially influenced and/or biased by student characteristics that they believe are predictive of academic success.

Earlier research has indicated that when judging students' performance, teachers are not only influenced by students' actual performance but also by other (and possibly unrelated) non-cognitive student characteristics such as engagement (Kaiser et al., 2013). However, and in addition to Kaiser et al. (2013), the second part of our study further shows that university teachers were often influenced by non-intellectual factors such as observed preparation and participation to group meetings, expressed motivation and interest, and personality characteristics such as extraversion and conscientiousness. In fact, approximately 68% of all statements concerned non-intellectual characteristics. Many of these student characteristics have been associated with actual academic performance in tertiary education, such as university teacher ratings of preparation and participation, conscientiousness, and motivational variables (e.g., Author, 2012a; Richardson et al., 2012). Although, university teachers were more likely to report student characteristics they perceived as positive for enhancing academic success, for the group of bachelor's students who eventually failed to complete the first year they reported a significantly larger proportion of negative statements. This finding can be seen as additional support for university teachers' relative judgment accuracy.

Even though many of the student characteristics that university teachers perceive as important have in reality been associated with actual academic achievement, it is possible that teachers are biased by their own beliefs. For example, in the current study, but also in the study by Furnham and Medhurst (1995), university teachers perceived extraversion as a positive predictor of academic achievement in higher education. However, a meta-analysis indicated that extraversion was unrelated to actual academic achievement (Poropat, 2009). Moreover, in an actual PBL environment, extraversion was negatively associated with first-year academic success (Author, 2012a). Therefore, although university teachers seem to be appreciative of outward going, socially skilled and assertive behaviors of students, the effectiveness of these behaviors for successfully completing the first year or bachelor's degree are not corroborated by actual research findings. The belief that certain student characteristics influence academic achievement, when they are in fact unrelated or differently associated, might lead to inaccurate judgments of students' performance. For example, Alvidrez and Weinstein (1999) found that teachers tended to overestimate a child's intelligence at age 4 when they perceived the child as independent, assertive, and interesting. In addition, Hinnant, O'Brien, and Ghazarian (2009) indicated that children's social skills were positively associated with teacher expectations for reading and math, indicating that teachers may overestimate the academic ability of students they find easy to manage during lessons. Future research should examine this further.

7.3 Limitations and Future Research

Our study revealed significant effects despite the relatively small sample size. However, future research could benefit from larger samples in order to examine interindividual differences in teachers' judgment accuracy. Indeed, previous research has identified individual difference with respect to teachers' ability to make judgments (e.g., Coladarci, 1986; Impara & Plake,

1998). It is believed that certain teacher characteristics might influence teacher ability to accurately judge students' performance, such as teaching experience or teaching philosophy. However, it is yet unknown which teacher characteristics influence judgment accuracy (Südkamp et al., 2012), which could be considered further.

According to the reflection-construction model, there are three ways interpersonal perception relate to reality, namely accuracy, self-fulfilling prophecy, and biases that influence or distort social perception (Jussim, 1991). Because earlier research has indicated that self-fulfilling prophecy effects are usually small and more likely to occur in the first two weeks of an school year (Jussim & Harber, 2005), we did not control for it in the present study. In the present study, teacher ratings were collected near the end of the first 5-week course. This was done to ensure that teachers had many opportunities to interact with the students in their groups to make accurate judgments (Funder, 2012; Jussim & Harber, 2005). In addition, our results demonstrated that only a small percent of the variance (2-5%) in actual academic success could be explained by between-teacher variance. Moreover, after each 5-week period students are randomly assigned to a new teacher group and teacher, making it unlikely for self-fulfilling prophecy effects to occur.

Further, it is unknown whether teachers can accurately judge the non-intellectual characteristics students actually possess, such as motivation and personality. In our study, this was not examined. We were interested in university teachers' perceptions of student characteristics (like personality) that teachers use to base their judgments regarding students' academic performances upon. However, we did not investigate how accurately student characteristics (like personality) were judged, because teachers' judgment accuracy are possibly more influenced by their own beliefs about students' characteristics instead of students' actual

characteristics (cf. Jussim, 1991). Although teacher judgments of students' academic performance are relatively accurate (Hoge & Colardarci, 1989; Südkamp et al., 2012), teachers are less accurate when they are asked to assess other achievement-related traits, such as engagement, competency beliefs, motivation, or school anxiety (e.g., Dicke, Lüdtke, Trautwein, Nagy, & Nagy, 2012; Gagné & St Père, 2001; Givvin, Stipek, Salmon, & MacGyvers, 2001; Kaiser et al., 2013; Urhahne, 2015).

Future research should examine whether the type of learning environment investigated (i.e., PBL) affects the accuracy and reasoning of teacher judgments. Accuracy of university teachers' judgments is of interest to a wide range of educators in higher education and has profound implications for the interview process and assessment of potential students. In addition, teachers' judgments may impact upon the type of support or engagement of educators with students in PBL environments. Whether this would produce a bias in teaching and learning would be a useful future study.

Finally, future studies should examine whether participation in professional development activities can promote judgment accuracy. A recent study by Thiede et al. (2015) indicated that participation in professional development improved the accuracy of teachers' monitoring of student learning. The professional development was designed to help teachers give mathematics instructions in a student-centered way and incorporate formative assessment into instruction. Professional development improved judgment accuracy as instructional practices were more focused on students' actual understanding than the content taught providing teachers with appropriate cues for judging student learning. Therefore, judgment accuracy of teachers might be improved by letting teachers focus more on conceptual understanding in group meetings. As mentioned, in PBL, teachers have multiple roles, such as guiding the learning process,

monitoring the group process, and providing feedback to students' individual learning process (Author, 2012b). Evaluation of students' behavior in discussion group meetings, now focuses mostly on the quality and quantity of engagement in class rather than on actual conceptual understanding. In future research, therefore, we could examine whether use of formative evaluation tools of conceptual understanding in addition to evaluation of engagement could promote judgment accuracy.

7.4 Conclusions and Implications

In sum, university teachers' chance ratings made early in the first bachelor's year were predictive of successful completion of the first year and the bachelor's program. In this study, we replicated previous findings regarding the asymmetrical effect of failure/success predictions by university teachers as our findings showed that it seems far more difficult to (accurately) predict students' failure than their academic success. Therefore, we need to consider when and why teachers are inaccurate in predicting academic achievement. In extending previous findings, we showed that student characteristics other than cognitive ability, such as participation and preparation, motivation and interest, and personality factors, often influence university teacher judgments of academic success. University teachers were most often influenced by observed learning activities in group meetings such as preparation and participation, which supports previous findings (Kaiser et al., 2013). Although many of these student characteristics have been associated with actual academic performance in higher education, it is still important to consider whether university teachers' judgments are biased. Student characteristics perceived to be important by teachers may very well be unrelated to actual achievement, but might nevertheless influence their judgments (Jussim, 1991). Moreover, factors like preparation and classroom participation might instigate the differential prediction of success versus failure, perhaps because

other (situational/personal) factors that induce failure are less observable or readily accessible to university teachers and/or might interact in more subtle ways with factors that are more likely to be observed by teachers (like preparation and classroom participation). Hence, failure in higher education (e.g., dropout) might be much more complex to predict than academic success due to a differential complexity, and therefore be an interesting avenue for further research, also given the high personal/economic costs of study dropout and failure in higher education.

Appendix A. Overview of coding scheme and examples

Student Characteristic	Valence	Examples
Demographic characteristics	Positive	-
	Neutral	"an older student"
	Negative	"too young"; "(-) a bit young"
Intellectual factors		
B.1. General cognitive ability	Positive	"intelligent"; "smart"; "understands the subject matter"
	Neutral	"I find it difficult to determine whether she understands everything"
	Negative	"does not understand the subject matter"; "seems to find it difficult"
B.2. Verbal ability	Positive	"passed the language test at once"; "I have read a paper written by this student and she possesses good writing skills"; "good verbal expression skills"
	Neutral	-
	Negative	"makes spelling errors"; Dutch language skills are poor"
B.3. Prior educational experiences	Positive	"was enrolled in bilingual secondary education (English and Dutch) which is very beneficial when compared to other students"; "has prior experiences with meetings and discussions"
	Neutral	"has finished vocational training, wants to do university"
	Negative	"already tried to pass this course 3 years ago but did not pass it"; "is enrolled in the first year for the second time"
Non-intellectual factors		
C.1. Personality – not specified	Positive	"personality is okay"
	Neutral	-
	Negative	-
C.2. Extraversion - Introversion	Positive	"takes initiative"; "social"; "oriented to others"
	Neutral	-
	Negative	"shy"; "timid"; "introvert"; "does not make contact with other students"
C.3. Conscientiousness	Positive	"determined"; "disciplined"; "well-organized"
	Neutral	-

	Negative	"lazy attitude"; "no discipline"
C.4. Collaboration readiness	Positive	"has interest in collaboration with other student"; "seeks help from other students"
	Neutral	-
	Negative	"does not listen to other students"
C.5. (In)stability	Positive	"self-confident"; "calm"; "sure of himself"
	Neutral	-
	Negative	"insecure"; "seeks confirmation"; "was so nervous during presentation, that she wanted to quit before it was finished"
C.6. Study priority	Positive	-
	Neutral	"has moved to Rotterdam and has joined a fraternity [and is committed to both leaving home and finishing his studies]"
	Negative	"busy social life which led to shortage in time"; "undertakes many activities in addition to studying"
C.7. Motivation and interest	Positive	"interested"; "motivated"; "enthusiastic"; "wants to learn"
	Neutral	"[Student had difficulty understanding concepts] it is unclear whether this is caused by lack of motivation"
	Negative	"does not seem motivated"; "does not show enthusiasm"
C.8. Preparation and participation	Positive	"well prepared"; actively contributes to discussions"; "can explain concepts in own words"; "asks critical questions" "can make connections"; "reads multiple resources"
	Neutral	-
	Negative	"always comes a few minutes late"; "not prepared"; "doubt whether she has the right study skills"
C.9. Other non-intellectual factors	Positive	"medication (+)"; "acceptation (+)" [both examples mentioned in the context of a psychological disorder]
	Neutral	"lives at home"; "has worked a couple of years before starting this course"
	Negative	"is naïve"; "has personal circumstances that can trouble study progress"

References

Author 2007 [blinded for peer review].

Author 2009 [blinded for peer review]

Author 2012a [blinded for peer review].

Author 2012b [blinded for peer review].

Author 2014a [blinded for peer review].

Author 2014b [blinded for peer review].

Abidin, R. R., & Robinson, L. L. (2002). Stress, biases, or professionalism: What drives teachers' referral judgments of students with challenging behaviors? *Journal of Emotional and Behavioral Disorders, 10,* 204-212. doi:10.1177/10634266020100040201

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91,* 731-746. doi:10.1037//0022-0663.91.4.731

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. In L. Wilkerson & W. H. Gijselaers (Eds.), *New Directions in Teaching and Learning: Issue 68. Bringing problem-based learning to higher education: Theory and practice* (pp. 3-12). San Francisco, CA: Jossey-Bass. doi:10.1002/tl.37219966804

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23,* 43-55. doi:10.1037/1045-3830.23.1.43

Central Bureau for Statistics. (2009). Jaarboek onderwijs in cijfers [Yearbook education in figures]. Voorburg: CBS.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37,* 319-338. doi:10.1016/S0092-6566(02)00578-0

Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78,* 141-146. doi:10.1037%2F0022-0663.78.2.141

Demaray, M. K. & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13,* 8-24. doi:10.1037/h0088969

Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences, 22,* 844-849. doi:10.1016/j.lindif.2012.04.004

Doherty, J., & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of non-cognitive factors that influence specific judgments. *Educational Studies, 11,* 41-60. doi:10.1080/0305569850110105

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*, 247-265. doi:10.1002/pits.20147

Educational Inspectorate. (2013). Onderwijsverslag 2011/2012. *[Educational report 2011/2012]*. Zoetermeer, the Netherlands: Educational Inspectorate. file:///E:/ov11-12-hfdst-5-hoger-onderwijs.pdf

Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the School, 35,* 163-172. doi:10.1002/(SICI)1520-6807(199804)35:2<163::AID-PITS8>3.0.CO;2-Q

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74,* 59-109. doi:10.3102/00346543074001059

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science, 21,* 177-182. doi:10.1177/0963721412445309

Furnham, A., & Medhurst, S. (1995). Personality correlates of academic seminar behaviour: A study of four instruments. *Personality and Individual Differences, 19,* 197-208. doi:10.1016/0191-8869(95)00026-3

Gagné, F., & St Père, F. (2001). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30,* 71-100. doi:10.1016/s0160-2896(01)00068-x

Gerber, M. M., & Semmel, M. J. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19,* 137-148. doi:10.1080/00461528409529290

Gijsel, M. A. R., Bosman, A. M. T., & Verhoeven, L. (2006). Kindergarten risk factors, cognitive factors, and teacher judgments as predictors of early reading in Dutch. *Journal of Learning Disabilities, 39,* 558-571. doi:10.1177/00222194060390060701

Givvin, K. B., Stipek, D. J., Salmon, J. M., & MacGyvers, V. L. (2001). In the eyes of the beholder: Students' and teachers' judgments of students' motivation. *Teaching and Teacher Education, 17,* 321-331. doi:10.1016/s0742-051x(00)00060-3

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school teachers. *Journal of Educational Psychology, 101,* 662-670. doi:10.1037/a0014306

Hoge, R. D., & Coladarci. T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59,* 297-313. doi:10.2307/1170184

Hovdhaugen. E. (2009). Transfer and dropout: Different forms of student departure in Norway. *Studies in Higher Education, 34,* 1-17. doi:10.1080/03075070802457009

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method on sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64,* 157-170. doi:10.1111/j.1467-9574.2009.00445.x

Impara, J. C., & Plake, B. S. (1998). Teachers' ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method. *Journal of Educational Measurement, 35,* 69-81. doi:10.1111/j.1745-3984.1998.tb00528.x

Jansen, E. P. W. A. (2004). The influence of the curriculum organization on study progress in higher education. *Higher Education, 47,* 44-435. doi:10.1023/B:HIGH.0000020868.39084.21

Jansen, E. P. W. A., & Bruinsma, M. (2005). Explaining achievement in higher education. *Educational Research and Evaluation, 11,* 235-252. doi:10.1080/13803610500101173

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57,* 469-480. doi:10.1037/0022-3514.57.3.469

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98,* 54-73. doi:10.1037//0022-3514.57.3.469

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9,* 131-155. doi:10.1207/s15327957pspr0902_3

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction, 28,* 73-84. doi:10.1016/j.learninstruc.2013.06.001

Kaufman, D. M., & Hansell, M. M. (1997). Can non-expert PBL tutors predict their students' achievement? An exploratory study. *Academic Medicine, 72*(Suppl. 1), S16-S18. doi:10.1097/00001888-199710001-00006

Muthén. L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Organisation for Economic Co-Operation and Development. (2013). *Education at a glance 2013: OECD indicators.* Paris: OECD Publishing. http://dx.doi.org/10.1787/eag-2013-en

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322-338. doi:10.1037/a0014996

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149-172). New York, NY: Springer. doi:10.1007/978-1-4614-2018-7_7

Richardson, M., Abraham, C., & Bond, C. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138,* 353-387. doi:10.1037/a0026838

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education, 30,* 175-194. doi:10.1080/0260293042000264262

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104,* 743-762. doi:10.1037/a0027627

Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33,* 200-210. doi:10.1177/002221940003300208

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., … Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education, 49,* 36-44. doi:10.1016/j.tate.2015.01.012

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago, IL: The University of Chicago Press.

Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. *The Review of Higher Education, 21,* 167-177.

Trouilloud, D. O., Sarrazin, P. G., Martinek, T. J., & Guillet, E. (2002). The influence of teacher expectations on student achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology, 32,* 591-607. doi:10.1002/ejsp.109

Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education, 45,* 73-82. doi:10.1016/j.tate.2014.09.006

Van den Berg, M. N., & Hofman, W. H. A. (2005). Student success in university education: A multi-measurement study of the impact of student faculty factors on study progress. *Higher Education, 5,* 413-446. doi:10.1007/s10734-004-6361-1

Van de Watering, G. A., & Claessens, S. (2003). Verschillen tussen de perceptie van tutoren op de leerprestatie en de toetsprestatie van hun studenten [Differences between tutors' perceptions of performance and students' actual achievement]. *Tijdschrift voor Hoger Onderwijs, 21,* 199-214.

Whitfield, C. F., & Xie, S. X. (2002). Correlation of problem-based learning facilitators' scores with student performance on written exams. *Advances in Health Sciences Education, 7,* 41-51. doi:10.1023/A:1014522513526

Table 1

Logistic Regression Models for First-Year Success and Obtaining the Bachelor's Degree in Time

				95% CI for OR	
	b	*SE*	OR	Lower	Upper
First-year success (0 = finished, 1 = failed)					
Threshold first year success	1.41*	0.57			
Chance rating of first year success	0.04***	0.01	1.04	1.02	1.06
Bachelor in time (0 = yes, 1 = no)					
Threshold bachelor in time	1.57*	0.77			
Chance rating bachelor in time	0.02*	0.01	1.02	1.001	1.04

Note. Thresholds are reported instead of intercepts, both are similar but have opposite signs. CI = confidence interval. OR = odds ratio. *p < 05, *** p < .001.

Table 2

Multinomial Logistic Regression Model for Successful Completion of the Bachelor's Program

				95% CI for OR	
	b	*SE*	OR	Lower	Upper
No degree vs. completion					
Threshold	2.31**	0.81			
Chance rating of bachelor's program success	-0.04***	0.01	0.96	0.95	0.98
Study delay vs. completion					
Threshold	1.37*	0.66			
Chance rating of bachelor's program success	-0.02**	0.01	0.98	0.96	0.99

Note. Completion of the bachelor's degree in time was used as the reference category. Thresholds are reported instead of intercepts, both are similar but have opposite signs. *p < 05, ** p < .01, *** p < .001.

Table 3

Classification Table First-Year and Bachelor's Program Success

	First study year		Bachelor in time	
	< 50% & > 50%	≤ 25% & ≥ 75%	< 50% & > 50%	≤ 25% & ≥ 75%
Hit	171	121	98	71
Miss	13	1	8	0
Correct rejection	13	5	14	6
False alarm	41	22	47	25

Table 4

Frequency of Teacher-Provided Reasons

Student Characteristics	Frequency			
	Negative	**Neutral**	**Positive**	**Total**
A. Demographic characteristics				
A.1. Demographic variables (i.e., age)	4	2	0	6
Total Category A	*4*	*2*	*0*	*6*
B. Intellectual factors				
B.1. General cognitive ability	30	3	119	152
B.2. Verbal ability	10	0	6	16
B.3. Prior education experiences	16	4	4	24
Total Category B	*56*	*7*	*129*	*192*
C. Non-intellectual factors				
C.1. Personality – not specified	0	0	10	10
C.2. Extraversion – Introversion	41	0	26	67
C.3. Conscientiousness	12	0	39	51
C.4. Collaboration readiness	5	0	16	21
C.5. (In)stability	22	0	6	28
C.6. Study priority	8	1	0	9
C.7. Motivation and interest	30	1	127	158
C.8. Preparation and participation	75	0	242	317
C.9. Other	12	3	2	17
Total Category C	*205*	*5*	*468*	*678*
Total of all categories	**265**	**14**	**597**	**876**

Table 5

Number and Percentages of Negative and Positive Teacher-Provided Reasons Calculated Separately for Successful and Unsuccessful University Students

	Successful students (n = 182, 659 statements)				Unsuccessful students (n = 56, 203 statements)			
	Negative		Positive		Negative		Positive	
A. Demographic characteristics	**3**	**0.46%**	**0**	**0%**	**1**	**0.49%**	**0**	**0%**
B. Intellectual factors	**37**	**5.61%**	**105**	**15.93%**	**19**	**9.36%**	**24**	**11.82%**
B.1. General cognitive ability	18	2.73%	96	14.57%	12	5.91%	23	11.33%
B.2. Verbal ability	8	1.21%	5	0.76%	2	0.99%	1	0.49%
B.3. Prior educational experiences	11	1.67%	4	0.61%	5	2.46%	0	0%
C. Non-intellectual factors	**120**	**18.21%**	**394**	**59.79%**	**85**	**41.87%**	**74**	**36.45%**
C.1. Personality – not specified	0	0%	8	1.21%	0	0%	2	0.99%
C.2. Extraversion – Introversion	29	4.40%	19	2.88%	12	5.91%	7	3.45%
C.3. Conscientiousness	6	0.91%	34	5.16%	6	2.96%	5	2.46%
C.4. Collaboration readiness	1	0.15%	14	2.12%	4	1.97%	2	0.99%
C.5. (In)stability	19	2.88%	6	0.91%	3	1.48%	0	0%
C.6. Study priority	5	0.76%	0	0%	3	1.48%	0	0%
C.7. Motivation and interest	18	2.73%	108	16.39%	12	5.91%	19	9.36%
C.8. Preparation and participation	34	5.16%	203	30.80%	41	20.20%	39	19.21%
C.9. Other	8	1.21%	2	0.30%	4	1.97%	0	0%
Total of all categories	**160**	**24.28%**	**499**	**75.72%**	**105**	**51.72%**	**98**	**48.28%**