Lost in translation : congruency of teacher and student perceptions of assessment as a predictor of intrinsic motivation in ethnodiverse classrooms

Pat El, R.J.

Lost in translation: Congruency of teacher and student perceptions of assessment as a predictor of intrinsic motivation in ethnodiverse classrooms

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 9 oktober 2012
te klokke 10.00 uur

door

Ron Jonathan Pat-El

geboren te Leidschendam, Nederland in 1980

Promotiecommissie
Promotor: Prof. Dr. P. Vedder
Promotor: Prof. Dr. M. Segers (Universiteit Maastricht)
co-Promotor: Dr. H Tillema

Prof. Dr. R. Bosker (Rijksuniversiteit Groningen)
Prof. Dr. P. van der Broek (Universiteit Leiden)
Prof. Dr. P. den Brok (Technische Universiteit Eindhoven)
Prof. Dr. C. Espin (Universiteit Leiden)

Contents

List of Tables

List of Figures

Acknowledgements

Van dit proefschrift heb ik veel geleerd, en ik ben daarvoor grote dank verschuldigd aan mijn promotoren Paul, Mien en Harm. Jullie begeleiding was erg fijn en jullie hebben op verschillende wijze mij op weg geholpen. Mien, bedankt voor je aanpak bij het vormen van een theoretisch raamwerk, en Harm voor je adviezen om definiëringen scherp te krijgen. Paul, bedankt voor je snelle, scherpe feedback op mijn geschreven werk, en dat je mij er steeds aan hebt herinnerd dat ik niet zo breedsprakig moet schrijven als dat ik praat. Het zal een voortdurend leerpunt blijven.

Ik wil mijn lieve ouders Paula en Eliahu bedanken voor al hun steun en inzet voor hun hele familie. Jullie hebben ieder op jullie eigen manier al jullie kinderen intrinsiek academisch gemotiveerd gekregen, en hard gevochten om ons op universiteiten te krijgen. Bedankt! Ik hoop dat jullie tevreden kunnen zijn met dit eindproduct waar jullie 31 jaar lang (voorbereidend) aan hebben bijgedragen. Ik wil ook mijn broers, Juval, Ido en Boaz bedanken voor alle adviezen, levendige discussies, gesprekken over hun vakgebieden, en natuurlijk de broodnodige digitale afleidingen de afgelopen jaren.

Ook jij bedankt Yael, voor je steun en je liefde die ik van je heb mogen ontvangen, ook wanneer ik overweldigd dreigde te raken. Ik houd veel van jou en hoop nog lang samen met jou nieuwe uitdagingen tegemoet te gaan.

Er zijn vele goede vrienden die bedankt moeten worden, maar in het bijzonder wil ik Christel, Merel en Coen bedanken die altijd enthousiast hebben meegedacht, of mijn monologen hebben aangehoord en Frederik voor zijn consulten en adviezen tijdens ons wekelijks sporten. And thank you Victoria for proofreading this dissertation.

Mijn collega's, Fatima, Vanya, Mitch, Jenny, Coen en Karin die het een plezier maakten om naar de faculteit te blijven komen, zeker als het tegenzat, en van wie ik veel wijsheden heb meegekregen. Ook jij bedankt Jan-Willem voor je regelmatige appeltijdgesprekken en hulp bij alle kleine doch belangrijke zaken die onverwachts bij een proefschrift komen kijken.

Speciale dank gaat ook uit naar Astrid en Sabine die zoveel hebben betekend voor de dataverzameling, maar ook voor de gezellige informele gesprekken tijdens de scriptiebegeleiding. Dank voor de vele studenten die via hun scripties aan de dataverzameling hebben bijgedragen.

1

General introduction

Student motivation

Fostering student motivation through assessment and instruction is generally considered important to achieve positive learning outcomes. Harlen and Crick's (2003) review of classroom motivation literature showed how important the focus on learning instead of grading is for students; when the focus of teachers is on helping students gain insight in their own learning processes students are more intrinsically motivated. Intrinsic motivation is the tendency to engage in activities for the inherent joy they bring. This is positively associated with persistence, mastery-learning goals, deep learning and well-being (Ryan & Deci, 2000) and is therefore worthwhile to foster in the classroom. How students generally become more intrinsically motivated through learning is well explained by Ryan and Deci's (2000) Self Determination Theory (SDT). SDT proposes that students have to satiate three universal 'needs' to feel motivated for an educational activity (Deci & Ryan, 1985; Ryan & Deci, 2000); (1) a feeling of being autonomous, (2) a sense of relatedness with others in the activity; and (3) experiencing the competence to fulfill a given activity. Sambell and McDowell (1998) showed that students who categorize their teacher as more of a 'teacher' than as a 'judge' experience more relatedness to their teacher, and students who do not feel free to learn in their own way report less experienced autonomy support and less intrinsic motivation for doing homework (Assor, Kaplan, & Roth, 2002). Negative evidence for the relationship between motivation and students' need to experience competence is provided by Black and Wiliam (1998a) who showed that feedback meant to foster student ability and feelings of competence given to low ability students may be perceived as proof of teacher's doubt about their competence.

The perceived learning environment plays an important role in explaining student motivation (Entwistle, 1991), and as such it is important that students recognize when teachers act to support their learning. However, research evidence is suggesting that teachers and students can strongly differ in their opinion of how their learning is supported by their teachers (e.g., Könings, 2007; MacLellan, 2001). It is difficult for students to accept feedback and support when they don't recognize it as such, and it is difficult for teachers to adapt when they don't experience that their instructional support is not effective or understood. It is not known how this lack of congruency between teachers' and students' perceptions of learning situations impacts student motivation. The goal of this dissertation is to investigate to what degree teachers and students agree on what is happening in the classroom, how possible dif-

ferences can be explained, and whether differences of perception are related to students' motivation for learning.

The formative assessment perspective: Assessments for Learning.

Since educational culture has shifted in the past decades from a testing culture to an assessment culture (Birenbaum, 2003), the importance of effective communication and understanding between teachers and students has become more salient. In a testing culture, instruction and assessment are seen as separate and independent entities (Segers, 2004) and 'assessment of learning' is the core purpose of assessment activities. This means that in a testing culture, assessments of performance (process) and achievement (products) are separated from their learning context and tend to focus primarily or exclusively on established or available knowledge in which test results are mainly collected to be used for summative purposes, such as grading. The current shift towards an assessment culture where assessments are part of the learning process ('Assessment for Learning') has transformed assessments into scaffolds aligned with instruction in support of learning. In order to help students become autonomous learners, much emphasis is put on facilitating student autonomy in their learning process. This means that assessments must be informative to that process, take place in more authentic contexts, and use a variety of measures to establish growth in achievements (Segers, 2004). Assessment then becomes part of the instruction and learning process as multiple assessment moments inform learners of their progress. Although a variety of modes of assessment are used to support student learning (e.g., peer assessment, portfolio assessment), all AfL practices have in common that they emphasize the role of students as active learners and the role of assessment as a tool for monitoring student progress in scaffolding learning. Assessment is seen to support learning when assessment information is both used to help students make progress in their learning efforts and when teachers adapt their teaching to meet the learning needs of students. Only when assessment information is used in this manner, does it gain its formative nature (P. Black & Wiliam, 1998b).

This shift from an Assessment *of* Learning to an Assessment *for* Learning perspective changed the role and position of teachers and students alike. Teachers stopped being passers on of knowledge and became mentors who motivate and inform students to demonstrate their learning, knowledge, and skills in diverse ways. Students changed from calculators of past achievements to self-regulated learners (Dochy & McDowell, 1997).

How AfL should be defined, and what its core principles are, has been explored by several researchers and institutions (e.g., OECD, 2005; Prosser & Trigwell, 1993; Stiggins, 2005). The principles proposed by the Assessment Reform Group (2002) summarize many of the positions taken in recent literature and have become well-accepted (Daugherty, Black, Ecclestone, James, & Newton, 2007). The central view of ARG's principles is that AfL is a process in which feedback is used as a means to bring students closer to their learning goals and which is connected to explicit criteria that are considered important by both the teachers and the students. The ten principles were first drafted to assist teachers in recognizing key dimensions of AfL-practices. Based on literature reviews and exchanges with assessment-experts the following principles

were formulated: AfL should (1) be part of effective planning of teaching and learning; (2) be recognized as central to classroom practice; (3) be recognized as key professional skills for teachers; (4) be sensitive and constructive; (5) take learner motivation into account; (6) promote commitment to learner goals; (7) promote a shared understanding of the criteria by which students are assessed; (8) include constructive guidance on how to improve; (9) increase learners' capacity to self-assess; and (10) recognize the full range of achievement of all learners.

In the eye of the beholders: teachers and students

The success of AfL is conditional on the successful integration of assessment into the learning process. This integration may be optimized by continual interaction between teachers and individual learners, in which feedback provision and its acceptance and utilization are key elements (P. Black & Wiliam, 2009). AfL is a two-way process in which not only students adapt their learning with information provided by assessments, but teachers need to adapt their teaching as well. In this respect it is of great importance that students' and teachers' perceptions on the nature and content of the assessment provided are congruent. If teachers believe they provide constructive feedback and communicate goals clearly but students do not recognize this, they are not likely to react to the support and its effectiveness in shaping student learning is diminished. In order for teacher assessment to feed into student learning the teacher may have to adapt word choice and complexity of the information entailed in the feedback to the students' capacities to understand the feedback. Learning involves detection and correction of errors (Argyris, 1977) and better learning comes from not only addressing actions to change outcomes, but from correcting errors in such a way that it involves the modification of teachers' underlying conceptions of what indicates good teaching (P. Black & Wiliam, 1998a; Assessment Reform Group, 2002). Unfortunately, it has become clear that teacher self-reports on their instruction are limited by a lack of reflective awareness (Keiny, 1994; Wubbels, Brekelmans, & Hooymayers, 1992), while student observations are shaped by their expectations, needs and ability to understand the instruction (Den Brok, Levy, Wubbels, & Rodriguez, 2003). Congruency in teacher and student assessment and learning related perceptions are important for the success of classroom interventional activities (Loughran, 2010). This notion leads to the hypothesis that the less congruent teachers and students are in perceiving AfL practices, the less motivating instruction and learning will be to students. Misaligned perceptions lead to misunderstandings between teachers and students about the meaning, usefulness and purpose of assessment information (e.g., Bartholomew, Parcel, Kok, & Gottlieb, 2001; Norman, 1986), and students can perceive the learning criteria to be implicit and 'hidden' while they are perceived as transparent by the teachers (Könings, 2007).

Teachers and students can hold opposite opinions on what is practiced in classrooms. For example, a study by MacLellan (2001) revealed that when the teachers believe they are not taxing their students, support them in their autonomy, and have specified clear learning objectives, students may have contradictory perceptions: a heavy workload, little room for autonomous decision-making, and uncertainty about the learning objectives. This mismatch can lead to frictions between teacher and student and possibly cause problems in

the classroom. Könings (2007) for example, has shown that students are less motivated, and learn more superficially whenever they perceive less teacher support, than their teachers report to give. Education is not an objective reality perceived similarly by all stakeholders.

Structure of the dissertation

The aim of this dissertation is to explore whether teachers and students are in agreement about the level to which AfL is practiced in the classroom, and whether perception congruencies help explain intrinsic motivation in the classroom. To test the hypotheses of this dissertation, several steps were taken, which are described in each successive chapter. Given the limited instrumentation available to compare student and teacher perceptions of the classrooms, an instrument was developed and validated, which is the content of Chapter 2. Chapter 3 is about the question to what degree perceptions of AfL are congruent between teachers and students, and whether these congruencies can be explained by teacher and student level variables. After testing the prevalence of incongruencies within classrooms, its association with intrinsic motivation, as mediated by SDT's basic needs is tested in Chapter 4. In Chapter 5 is about the question whether the proposed relationship between AfL perception congruencies and motivation can be generalized beyond diverse ethnic groups in the Netherlands.

Chapter 2: A reliable and unbiased instrument is needed when comparing student and teacher reported perceptions of AfL-practice, however such measures are scarce. To validly test the dissertation's central hypothesis it is important that perceptions of students and teacher can actually be compared in terms of similarity of constructs measured, and also in terms of interpretation of the outcomes of the measurement instrument. In Chapter 2, the AfL construct is defined, and used as the basis for the development of a self-report questionnaire in a sample of 1422 students and 237 teachers. The aim is to develop a self-report tool, called the Assessment for Learning Questionnaire (AfL-Q) that can be used to validly compare teacher and student perceptions in order to answer the research questions posed in this dissertation. The data for this study were collected in November and December 2007.

Chapter 3: The argument is made that congruent perceptions of AfL are important for its effectiveness, yet little is known about the degree to which student and teacher perceptions are congruent and what factors would contribute to perception congruency. This chapter's focus is on the question to what extent teachers and students hold mismatched perceptions of AfL-practice. Explanations are sought from the perspective of self-verification theory, which predicts that efficacious teachers are more likely to be incongruent with their students perceptions, and from the perspective of language proficiency, which predicts that students need to have a strong grasp of the language of instruction in order to recognize and understand feedback. A sample of 650 students and 38 teachers was used to test the hypotheses. The aim of this chapter is to evaluate student and teacher perceptions of their AfL practice and find factors that may explain possible individual variance between students and teachers. The data for this study were collected in the period between April and June 2009.

Chapter 4: The central model of the dissertation will be tested in Chapter 4 for empirical evidence of the hypothesis that incongruencies are detrimental to student motivation, and that this can be explained through the basic needs as proposed in the Self-determination theory of motivation (Ryan & Deci, 2000), namely the needs of feeling competent, autonomous, and related to their teacher. The significant predictors of congruency found in Chapter 3 are also included in the model to test whether these predictors can help explain the tested relationship between congruency and motivation. The sample used to test this model consisted of 1466 students and 89 teachers. The data were collected in the Spring of 2010 and the Fall of 2010.

Chapter 5: Having established in Chapter 4 how teacher-student perception differences relate to students' intrinsic motivation, this relationship is further explored in Chapter 5. Even though SDT proposes that the link between intrinsic motivation and basic need fulfilment is universal (Chirkov, 2009), whether the association between the perceived learning environment and motivation is universal is still debated. One consistent finding in educational research is that immigrant students (especially from the second generation) are usually more intrinsically motivated for learning than national contemporaries. The question posed in this Chapter is whether ethnic differences in motivation can be explained by ethnic differences in teacher-student congruency, or whether the proposed relationships in Chapter 4 have differential effects for different ethnic groups. To this end, a sample of 775 students and 58 teachers was taken from the sample used in Chapter 4. Classrooms were included in this sample if they included, in addition to Dutch students, at least one student from the largest ethnic minority groups in the total sample, namely second generation Moroccans and Turks.

Chapter 6: The final chapter summarizes the findings, discusses possible theoretical and practical implications and points at some methodological challenges and desires for future research.

2

Validation of Assessment for Learning Questionnaires for Teachers and Students[1]

Abstract

Background: Assessment can be a powerful force in promoting student learning. Still, few measures exist to gauge Assessment for Learning in the classroom. Literature on AfL suggests that it encompasses both a monitor to track student progress as well as a scaffold to show or help students recognize in what areas they need to improve.
Aims: Based on a review of recent attempts to measure the Assessment for Learning (AfL) we constructed Assessment for Learning Questionnaires for Teachers (TAFL-Q) and for Students (SAFL-Q) for evaluating perceptions regarding AfL practices in classrooms using matching items.
Sample: The total sample included 1,422 students (49% girls, 51% boys) and 237 teachers (43% females, 57% males) in lower vocational secondary education.
Methods: The 28-item questionnaires were examined by means of confirmatory factor analysis using EQS on one random half of the sample. The CFA was cross-validated on the second half. Measurement invariance tests were conducted to compare the student and teacher versions of the questionnaires.
Results: Confirmatory Factor Analysis revealed a stable second-order two-factor structure which was cross-validated: perceived monitoring, and perceived scaffolding subsumed under a common factor: AfL. Tests for measurement invariance showed that the parallel constructs were measured similarly for both students and teachers.
Conclusion: The TAFL-Q and SAFL-Q capture the construct AfL in two subscales: Monitoring and Scaffolding, and allows for comparisons between teacher and student perceptions. The instruments can be a useful tool for teachers and students alike to identify and scrutinize assessment practices in classroom.

[1]This chapter was published as: Pat-El, R.J., Tillema, H., Segers, M.S.R., & Vedder, P. (2011). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology.*

2.1 Introduction

Assessment for Learning

Current research into assessment as a tool to support student learning is increasingly focused on how this support is perceived. Studies focus either on the question of 'how' it is perceived or on 'how much' of it is perceived by students or instructors. There is evidence emerging that teachers and students differ in their perception of both the extent to which assessment is integrated in instruction as well as its content. However, quantitative instruments to properly compare student and teacher perceptions are still lacking.

It is generally accepted that assessment is a powerful tool to inform and support student learning. In this respect, the concept 'Assessment for Learning' (AfL) has been introduced to counterbalance the majority of the attention paid to 'Assessment of Learning', or performance assessment, in classrooms and schools (e.g., P. Black & Wiliam, 1998a; Crooks, 2001; Stiggins, 2005). In the past decades, several models have been proposed that integrate summative testing within the learning processes in formative ways; such as formative evaluations (Scriven, 2002) or informal feedback (Boud, 2000). Assessments have become a critical link between curriculum, teaching, and learning (National Council of National Council of Teachers of Mathematics, 1995). Central to AfL is the integration of assessment into instruction as an ongoing process, where assessment information is used by teachers to adjust their teaching and by students to adjust their learning processes (Popham, 2008). The definition of assessment itself has become very broad and an umbrella term for any information that can help teachers gain better insights into students' understanding. The notion of AfL integrates assessments of performance (process) and achievement (products) within the learning context (Segers, 2004). AfL is characteristic of an assessment culture, that builds on scaffolding, aligning instruction in support of learning, and providing students with greater autonomy in their learning processes (P. Black & Wiliam, 1998a). The AfL perspective revolves around repeated, varied and often informal assessments which are informative to the learning process (Wiggins, 1993), and take place in authentic contexts (Darling-Hammond, 2010).

Even though different scholars emphasize slightly different AfL principles (for example, Assessment Reform Group, 2002; P. Black & Wiliam, 1998a; James & Pedder, 2006), our focus here is on two factors, viz; a monitor to track student progress and a scaffold to show or help students recognize in what areas they need to improve (Stiggins, 2005). The importance of these two factors is stressed and clarified in the ARG (2002) guidelines as well as by the Council of Chief State School Officers (CCSSO) (Darling-Hammond, 2010) and were explored in several grounding reports and studies (e.g., OECD, 2005; Prosser & Trigwell, 1993).

Although there is a substantial body of quantitative research into effects of AfL on learning, the questionnaires used, especially those pertaining to the perception of practices of AfL as a tool to evaluate the promotion of learning through assessment, have rarely received detailed psychometric attention. This is the first challenge to be tackled in the current study.

Perceptions of AfL

In order to understand how assessment influences student learning, research has addressed conceptual attitudes towards assessment practices (Brown, 2008) and perceived/observed practice (Gibbs & Simpson, 2003). Hirschfeld and Brown (2009) showed that in secondary, as well as in higher education, general conceptions of assessment purposes affect student attainment and well-being. Few studies have evidenced the influence of students' perceptions of applied assessment practices on their approach to learning (Entwistle, 1991; Struyven, Dochy, & Janssens, 2005). Teachers' perception of the assessment practice has received even less attention and the scant studies available show that teachers' and students' perceptions have very little overlap (Schultz, 2001). Such overlap, or alignment, may be difficult to achieve. Due to their different roles and expertise, students and teachers are likely to differ in the opportunities and focus of assessment. Könings (2007) suggests that teachers are more alert than their students to the perception of clarity of learning goals, motivating subject matter, and engaging learning environments for learning. Misalignment in the perceptions of assessment practices between teachers and students is problematic because it is likely to lead to misunderstanding and misinterpretation of the assessment information, its meaning, and purpose (Bartholomew et al., 2001). MacLellan (2001) and Loughran (2010) showed that in order to facilitate student learning, teachers need to make sure that at least teachers' and students' perception of assessment goals are in alignment. When the interactions between teachers and students are characterized by a misalignment, the difficulties involved in compensating or correcting this issue remain unclear. To facilitate early detection or to make sure that teachers' and students' perception of assessment situations and interactions are aligned it is indispensable to measure both teacher and student perceptions of AfL. Given these studies and the examples and arguments they present, we argue that it is important to distinguish between and measure both teacher and student perceptions of AfL practices in order to optimize the effects of assessment on student learning.

Measuring Perceptions of Assessment for learning

Current instruments are, in our opinion, not adequate for the designed objectives. They either show a number of methodological shortcomings, such as lacking construct validity and low internal consistency of scales (e.g., Gibbs & Simpson, 2003), or do not allow the matching of student and teacher perceptions (e.g., MacLellan, 2001). For example, the Assessment Experience Questionnaire (Gibbs & Simpson, 2003) is a Likert-type questionnaire for use in higher education. It measures a variety of AfL constructs such as quantity, timing and quality of feedback in addition to students' study efforts as well as how students perceive the examination. However, this instrument showed low internal consistencies. Another example regards the questionnaire developed by James and Pedder (2006), which was constructed to measure primary and secondary school teachers' beliefs and appraisals of AfL practices using Likert-type responses. The instrument showed low factor loadings. Furthermore the discrimination between measured AfL constructs remains unclear. One can see that measuring congruency in perception between teachers and students has proven difficult.

From a measurement perspective, structural confirmations of the dimensionality of AfL constructs have not been tested. The key principles of AfL have not yet been captured instrumentally, nor were they quantitatively validated. Available instruments on AfL practices do not comprehensively capture how students and teachers perceive the interrelated AfL principles that are applied in their classrooms. This study aims to develop and structurally validate a questionnaire designed to measure student and the teacher perceptions of AfL and which further allows direct comparisons of teacher and student perception of their AfL practice.

2.2 Method

Sample

The target population consisted of students and teachers in lower vocational education in the Netherlands. This educational track includes approximately 60% of the Dutch secondary school students (Central Bureau of Statistics, 2007). Students in the upper secondary educational system, those preparing for higher education, have a different pedagogical approach. This sample was chosen because review studies such as those from Black and Wiliam (1998a) indicate that the highest learning gains from using AfL are likely to be observed in the lower tracks. The individual students and teachers were the units of analysis. The questionnaire was administered in 80 schools to 1422 students (49% girls, 51% boys). From each school, three teachers were selected along with a random selection of six students from these classes. Teachers were selected from all disciplines, ranging from arts to sciences. The median student age was 14 years old (*range* = 11 - 18). We sampled 237 teachers (43% females, 57% males) who had a median age of 43 years old (*range* = 22 - 65). The classroom response rate was 98.75%. On average, the teachers had been active in the field of teaching for 15.3 years (SD = 11.11) and active at their current school for 8.8 years (SD = 8.27). Male teachers (Med = 45, range = 23 - 65) were slightly older than female teachers (Med = 37, range = 22 - 63, U = 5092.00, p = .001). The participating schools were spread across different regions in the Netherlands to avoid a bias towards urban areas.

Instrument

Student Assessment for Learning Questionnaire (SAFL-Q) and the Teacher Assessment for Learning Questionnaire (TAFL-Q) consisted of 28 items divided into two scales: (1) perceived monitoring (16 items); (2) perceived scaffolding (12 items). The items were measured on a five-point Likert scale (1 = strongly disagree, and 5 = strongly agree) (see Table 2.2 and Appendix for items).

Procedure

Schools were visited by research assistants who distributed the questionnaires in classrooms. Prior to administering the SAFL-Q and the TAFL-Q the purpose of the study was explained to teachers. The questionnaires started with a short instruction explaining the purpose of the study, presenting some examples on how to work with Likert response scales and directing teachers' and students'

attention to actual assessment practices in their current class. Completion took about 25 minutes.

Data analysis

Because of the large number of respondents needed to validate the questionnaire, confirmatory analyses were only conducted on the SAFL-Q. The validation was conducted in three steps: (1) Confirmatory analysis to determine the robustness of the factor structure on one randomly determined half of the dataset (training); (2) cross-validation using multi-group invariance testing between the first half of the dataset and the unused second half of the dataset; and (3) tests of multi-group invariance to determine whether the teacher and student questionnaires measure parallel constructs similarly.

Confirmatory factor analysis (CFA).

To test the stability of the explored factor structure, CFA was performed on the first data set in structural equation modelling software: EQS version 6.1. To interpret a model's fit, the following indicators were used: Root Mean Squared Error of Approximation (RMSEA) and Standardized Root Mean Residual (SRMR) below 0.05 and Comparative Fit Index (CFI) scores above 0.95 indicate good fit (Browne & Cudeck, 1992) and RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). Because the χ^2 statistic becomes increasingly unreliable in sample sizes > 250 (Byrne, 2006), a slightly less sample-dependent statistic was used: χ^2 divided by its degrees of freedom (df) (Bollen, 1989). CFI is penalized for model complexity, which means that in complex models, CFI might drop. A measure that does not penalize large or complex models such as CFI does, is the Gamma (γ) (Fan & Sivo, 2007) which is calculated based on the number of manifest variables, df, and RMSEA, and should have values above .90 for acceptable fit and above .95 for good fit.

Four types of models were tested: (a) a uni-dimensional model; (b) an uncorrelated-factor model; (c) a correlated-factor model; and (d) a second order model with latent factors explained by a common factor.

Cross-validation.

To test the robustness of the explored factor structure, CFA performed on the training set was repeated on the test set and their invariance were tested using multi-group invariance testing.

Teacher-student invariance.

Increasing levels of equality constraints were applied to the tested models to assess structural invariance of the questionnaires between teachers and students. The tested types of invariance are (1) dimensional: are the same number of common factors present in each group?; (2) configural: are common factors associated with the same items across groups?; (3) metric (invariant first and second order factor loadings): do common factors have the same meanings across groups?; (4) strong factorial (invariant variable and factor intercepts):

are comparisons of group means meaningful?; (5) strict factorial (invariant error terms): are comparisons of group means and observed variables defensible (Gregorich, 2006)? Strict factorial invariance has been argued as too strict of a criterion (Byrne, 2006) and is generally not seen as a requirement for invariance between populations (Wu, Li, & Zumbo, 2007). A critical value of -0.01 ΔCFI will be used to judge invariance (Cheung & Rensvold, 2002).

2.3 Results

Confirmatory Factor Analysis

Confirmatory factor analysis was used (EQS version 6.1) to test the stability of the two-component structure of the 28 item SAFL-Q. No Heywood cases - negative estimations of error variances for an indicator of a latent variable - were found in any of the fitted models. A one-factor model was fitted (model A) which showed a weak fit (χ^2/df = 3.61; CFI = .85; γ = .92; SRMR = .06; RMSEA = .06). This indicated that monitoring and scaffolding were correlated, but distinct factors. An uncorrelated two-factorial model with added factor-correlations was fitted (model B) which showed a weak fit (χ^2/df = 3.44; CFI = .86; γ = .92; SRMR = .17; RMSEA = .06). A two-factor model with added factor-correlations (model C) showed a significantly improved and good fit ($\Delta\chi^2(1) = 396.68$, $p < .001$; CFI = .92; γ = .96; SRMR = .04; RMSEA = .04). Finally, a second-order two-factorial model with a common AfL factor was fitted (model D) which showed an equally good fit ($\Delta\chi^2(2) = 0.00$; $p = 1.00$; γ = .96; SRMR = .04; RMSEA = .04) which indicated that the two distinct factors monitoring and scaffolding were part of a single underlying construct. Table 2.1 summarizes the fit of the tested models. Table 2.2 shows the factor loadings of the common factor CFA.

Table 2.2: Second order two-factor CFA standardized factor loadings on 28-items and second-order factor loadings for Monitoring and Scaffolding with their common AfL-factor (loadings < .30 are suppressed).

	CFA	**Uniqueness**
Item	**monitor scaffold**	
1. My teacher encourages me to reflect on how I can improve my assignments	.56	.83
2. After examining my test results, my teacher discusses the answers I gave to the test with me	.39	.92
3. Whilst working on my assignments, my teacher asks me how I think I am doing	.53	.85
4. My teachers allow me to think about what I want to learn in school	.50	.86
5. My teacher gives me the opportunity to decide on my own learning objectives	.43	.90

Table 2.1: Fit indices of the SAFL-Q factor models.

Model	χ^2	**df**	χ^2**/df**	**SRMR**	**CFI**	γ	**RMSEA**	**90% CI**	**Model Comparison**	$\Delta\chi^2$**(df)**
Model A: 1-factor	1269.33	351	3.61	.055	.845	.917	.060	.056, .064	—	
Model B: 2-factor; uncorrelated	1204.67	350	3.44	.167	.856	.922	.058	.054, .062	—	
Model C: 2-factors; correlated	807.98	349	2.32	.039	.923	.956	.043	.039, .046	C vs B	396.68**
Model D: 2-factors; with one second order common factor	807.98	347	2.33	.039	.922	.956	.043	.039, .047	D vs C	0.00

** = $p < .00$

6. My teachers inquire what went well and what went badly in my work	.62		.79
7. My teacher encourages me to reflect on my learning process and to think about how to improve next time	.67		.75
8. My teacher stresses my strengths concerning learning	.53		.85
9. My teacher identifies my weaknesses concerning learning	.37		.93
10. I am encouraged by my teacher to improve my learning process	.61		.80
11. My teacher gives me guidance to assist my learning	.60		.83
12. My teacher discusses assignments with me to help me understand the subject matter better	.56		.80
13. My teacher discusses with me the progress I make	.61		.78
14. After each assessment my teacher informs me how to improve the next time	.64		.77
15. My teacher discusses with me how to exploit my strengths to improve my assignment	.70		.71
16. My teacher and I consider ways to improve my weak points	.73		.69
17. When I do not understand a topic, my teacher tries to explain it in a different way		.58	.82
18. My teacher provides me with hints to help understand the subject matter		.65	.76
19. During class I have an opportunity to show what I have learned		.49	.87
20. My teacher asks questions in a way I understand		.67	.74
21. My teacher asks questions that help me gain understanding of the subject matter		.70	.72
22. My teacher allows for my contribution during the lesson		.56	.83
23. I have the opportunity to ask my classmates questions during the lesson		.43	.90
24. My teacher makes me aware of the areas I need to work on to improve my results		.44	.90
25. There is an opportunity to ask questions during the lesson		.50	.87
26. I am aware of the criteria by which my assignment will be evaluated		.54	.85
27. When I receive an assignment it is clear to me what I can learn from it		.62	.78
28. My assignments allow me to show what I am capable of		.47	.89
		AfL-common factor	
F1. Monitoring		.85	.53
F2. Scaffolding		.88	.47
Pearson r Monitoring - Scaffolding = .63, $p < .001$			

Note: The translation is based on the Dutch original validated in this study

Internal Consistency

Cronbach's α of the two subscales in the teacher and the student questionnaire provided convincing support for the reliability of the subscales. Table 2.3 shows αs and scale characteristics. The mean scores of teacher perceptions

were almost a point higher than student perception scores.

Table 2.3: Reliabilities and scale statistics

	Scale	Cronbach's α	N	Mean	SD
TAFL-Q	Monitoring	.87	237	3.70	0.54
	Scaffolding	.77	237	4.35	0.38
SAFL-Q	Monitoring	.89	1244	3.06	0.74
	Scaffolding	.83	1244	3.82	0.64

Cross-Validation

The training ($N_1 = 726$) and the test ($N_2 = 696$) sample subset were compared to test the robustness of the two-factor model. Multi-group comparisons in EQS, with varying levels of equality constraints, confirmed that the two-factor structure on the 28-items was equally robust in both random sets (Table 2.4), as the two randomly sampled halves were strictly factorially invariant. The baseline hypothesized model in the test sample (χ^2/df = 2.33 CFI = .92; γ = .96; RMSEA = .04) was confirmed in the training sample (χ^2/df = 2.28; CFI = .92; γ = .96; RMSEA = .04). The two samples proved strictly factorially invariant ($\Delta CFI < .001$; χ^2/df = 2.54; CFI = .90; γ = .94; RMSEA = .047).

Teacher-student invariance

Structural equivalence of the two-factor structure between students ($N = 1244$) and teachers ($N = 237$) was tested in EQS by applying increasingly restrictive equality constraints. All equality constraints were correctly imposed. An overview of the steps and results is given in Table 2.5. Testing of the baseline hypothesized model for the full sample of students ($N = 1422$) yielded a good fit to the data (χ^2/df = 3.55; CFI = .93; γ = .96; RMSEA = .04). The Langrange Multipliers Test statistics, which help identify whether additional parameters should be added as well as Wald statistics, which help identify whether parameters should be removed, revealed that the fit for the student baseline model remained consistent with the hypothesized model. Testing for the validity of the TAFL-Q ($N = 237$) yielded a weaker, yet still acceptable fit to the data (χ^2/df = 2.41; CFI = .74; γ = .86; RMSEA = .08). Thus, with the exception of CFI, the fit measures indicate that the teacher baseline model is also consistent with the hypothesized model.

The common factor model was found to be strongly factorially invariant (Model 5). To prevent under-identification (more estimated parameters than observed) in the testing of the invariance of latent factor intercepts, the fixed intercepts were constrained to their observed intercepts values found in a separate model in which first- and second-order factor loadings were constrained equal across groups, however the observed variable intercepts were freely estimated. These values were .734 and .646 respectively. Testing revealed a good-fitting model (χ^2/df = 4.23; CFI = .93; γ = .93; RMSEA = .05). Even though there

Table 2.4: Fit indices of multi-group comparisons between training and testing sample of the student dataset of the second order two-factor model

Model	χ^2	**df**	$\chi 2$/**df**	**CFI**	γ	**RMSEA**	**90% CI**	**Model Comparison**	**ΔCFI**
Model 1: Configural	1847.62	694	2.66	.904	.946	.048	.046, .051	—	
Model 2: First-order factor loadings invariant	1871.95	720	2.60	.904	.942	.049	.045, .050	2 vs 1	<.001
Model 3: First- and second-order factor loadings invariant	1871.95	722	2.60	.904	.942	.049	.045, .050	3 vs 1	<.001
Model 4: Strict Factorial: Error variances invariant	1893.45	746	2.54	.904	.944	.047	.044, .049	4 vs 1	<.001

was a substantial change in CFI, it was in the direction of improved fit. Testing for the strict factorial invariance by constraining all estimated error terms resulted in a less than acceptable fit (χ^2/df = 5.36; CFI = .89; γ = .89; RMSEA = .07). The strong reduction in fit between model 6 and model 1, combined with the Δ CFI of .013, indicates that the TAFL-Q and the SAFL-Q are likely not strictly invariant.

2.4 Discussion

The aims of this study were to construct and validate a self-report questionnaire measuring teacher and student perceptions of 'Assessment for Learning' (AfL) practices in their classroom. Based on a review of the literature on AfL and drawing on the principles outlined by ARG and CCSSO, we developed two instruments to gauge AfL: the SAFL-Q, for students and the TAFL-Q for teachers. As expected, the results of our analyses show a good fit for a two-factor solution with 28 items that was robust for both the SAFL-Q and the TAFL-Q. The two-factors in the student and teacher questionnaires were labelled: 'monitoring' , and 'scaffolding'.

Strictly speaking, correlations showed that the two factors in both the SAFL-Q and the TAFL-Q are not independent. This is not surprising given that both practices in AfL, i.e., monitoring and scaffolding, are by their very nature linked. The empirically established two factors in this study clearly represent the constructs found in the literature. The factor 'scaffolding' has items that refer to clarification of learning goals and criteria and to classroom questioning, which are largely instruction-related processes. The factor 'monitoring' has items that deal with feedback and self-monitoring which share a common intent to optimize learning. However, the overlap is not complete. One reason might be that, although the important building blocks of AfL are relatively well-defined (Assessment Reform Group, 2002; Stiggins, 2005), they are not necessarily mutually exclusive or orthogonal. Nevertheless, the two factors cover many of the conceptually stated principles of AfL. We opted for an open approach of Assessment for Learning. It is not the assessment itself which determines whether assessment is summative or formative in nature (function), but rather how the assessment information is used (goal) (P. Black & Wiliam, 1998b). This means that feedback, even when summative in character, can become formative when its information is used to inform students how, and in what way, their performance falls short of a goal, and how to improve (Kluger & DeNisi, 1996). Although there is an ongoing debate as to which exact ingredients define assessment for learning (Wiliam, 2011), this study adds evidence to the suggestion (Stiggins, 2005; Pat-El, Segers, Tillema, & Vedder, 2008) that the reduction of dimensionality of the AfL construct can lead to more parsimonious overarching clusters, such as suggested by James and Pedder (2006, p. 110). This suggestion finds further basis on the works of Gipps (1994) in which AfL is called distinctive by its focus in particular on "learning as an explicit focus of classroom activity and students' learning autonomy", which we have summarized as activities to monitor student progress, and scaffolding activities to realize growth.

Another aim of our study was to establish questionnaires that would mea-

Table 2.5: Fit measures of multi-group comparisons between teachers and students

Model	χ^2	**df**	χ^2**/df**	**CFI**	γ	**RMSEA**	**90% CI**	**Model Comparison**	**ΔCFI**
Model 1: Configural; No constraints	2095.06	694	3.02	.898	.944	.049	.047, .052	—	
Model 2: First-order factor loadings invariant	2236.28	720	3.11	.890	.940	.050	.048, .053	2 vs 1	.008
Model 3: First- and second-order factor loadings invariant	2236.28	722	3.10	.890	.939	.050	.048, .053	3 vs 1	.008
Model 4: First- and second-order factor loadings; observed variable intercepts invariant	3266.88	752	4.34	.896	.927	.054	.052, .056	4 vs 1	.002
Model 5: First- and second-order factor loadings; observed variable and latent factor intercepts invariant	3266.94	752	4.23	.925	.927	.054	.052, .056	5 vs 1	-.027
Model 6: First- and second-order factor loadings; observed variable and latent factor intercepts; error variances invariant	4189.99	778	5.39	.885	.892	.066	.064, .069	6 vs 1	.013

sure the perceptions of AfL practices in classrooms and could detect (possibly differing) perceptions of students and teachers. Previously available self-report research tools have not convincingly compared evaluative perceptions of actual practices of both populations –teacher and students- that are specifically related to promoting learning (MacLellan, 2001). The questionnaires in the current study used matching items and the measurement invariance between the two questionnaire versions indicate that the two-factor structure is not only robust, but can be similarly interpreted for both teachers and students alike. Our findings that corresponding latent factor intercepts are invariant across groups suggests that group differences in estimated factor means are unbiased and group differences in observed means are directly related to group differences in factor means and not contaminated by differential response bias (Gregorich, 2006). This means that, when used to compare perceptions of both student and teacher populations, differing perception scores can be attributed to different perceptions and not to different interpretation of the factors. The implication for researchers is that mean scores of teachers and students can directly be compared and difference scores can be used to measure the degree to which teachers' and their students' views about the perceived same classroom practices match without bias. Due to observations that students can perceive, for example, learning criteria to be more implicit and 'hidden' than their teachers believe they are, comparing student and teacher perceptions of AfL has become the focus of research trying to optimize the effects of assessment on student learning (Könings, 2007). If AfL is to integrate assessment with learning, it has to be a process of continual interaction between teachers and individual learners, in which feedback provision and its acceptance and utilization are key elements (P. Black & Wiliam, 2009). In this respect, it is of great importance that students' and teachers' perceptions on the nature and content of the assessment provided are congruent. Teachers may have to adapt word choice and complexity of the feedback-information to help students understand the feedback (P. H. Vedder, 1985). Conversely, it is important that students anticipate teacher instructions and feedback as personal expectations. In this manner, congruency in teacher and student assessment and learning related perceptions are important for the success of classroom interventions (Bartholomew et al., 2001; Loughran, 2010; Norman, 1986). Seeking discrepancies between student and teacher perceptions of AfL can help detect why some instructional activities to foster learning sometimes have unexpected or contradictory effects (Olkinuora & Salonen, 1992).

The instrument(s) in this study are more context-sensitive than overall measures of assessment conceptions as the questionnaires are designed to elicit responses on actual assessment practices closely related to the (promotion of) learning going on in classrooms. The results indicate much higher means in AfL perceptions for teachers over students. Reasons for this average misalignment could either be due to teachers being more knowledgeable about, or rather students being less aware of, the intent and purpose of their assessment. Students' perceptions in this sample are nested within classrooms and as such global comparisons of perception means will be biased. Comparisons that account for the multilevel nature of the relationships between specific teachers and students are needed.

Confirming the robustness of the outcomes in additional populations or different types of education would enhance the usefulness of the SAFL-Q and

the TAFL-Q. The population in this sample (lower vocational secondary education) can generalize to lower track students, who are being prepared for vocational education, however the sample cannot provide evidence for generalizability to students in higher tracks, who are prepared for higher education. Confirming invariance in the entirety of the Dutch secondary education would indicate its usefulness in broader contexts, such as countries where secondary education is less structurally diversified. At present we believe the instrument is a valid representation of AfL principles and able to gauge how assessment can contribute to learning. Based on our results we argue that the instruments can be useful tools for teachers and students alike in identifying and scrutinizing assessment practices in the classroom and eventually be supportive in establishing a common and shared assessment approach to scaffold learning in classrooms.

2.5 Appendix A. TAFL-Q items and subscales English.

Monitoring

1. I encourage my students to reflect upon how they can improve their assignments
2. After a test I discuss the given answers with each student
3. While working on their assignments I ask my students how they think they are doing
4. I involve my students in thinking about how they want to learn at school
5. I give my students the opportunity to decide on their learning objectives
6. I ask my students to indicate what went well and what went badly concerning their assignments
7. I encourage students to reflect upon their learning processes and how to improve their learning
8. I inform my students' on their strong points concerning learning
9. I inform my students' on their weak points concerning learning
10. I encourage my students to improve on their learning processes
11. I give students guidance and assistance in their learning
12. I discuss assignments with my students to help them understand the content better
13. I discuss with my students the progress they have made
14. After an assessment I inform my students on how to improve their weak points
15. I discuss with my students how to utilize their strengths to improve on their assignment
16. Together with my students I consider ways on how to improve on their weak points

Scaffolding

1. I adjust my instruction whenever I notice that my students do not understand a topic
2. I provide my students with guidance to help them gain understanding of the content taught
3. During my class, students are given the opportunity to show what they have learned
4. I ask questions in a way my students understand

5. By asking questions during class I help my students gain understanding of the content taught
6. I am open to student contribution in my class
7. I engage my students by discussing answers to assignments during class
8. I ensure that my students know what areas they need to work on in order to improve their results
9. I give my students opportunities to ask questions
10. My students know what the evaluation criteria for their work are
11. I ensure that my students know what they can learn from their assignments
12. I can recognize when my students reach their learning goals

Note. This English translation is based on the Dutch original validated in this study.

2.6 Appendix B. Items and subscales of the TAFL-Q (Dutch)

Monitoring

1. Ik stimuleer mijn leerlingen om na te denken over hoe zij hun schoolwerk kunnen verbeteren.
2. Na het nakijken van een toets bespreek ik met iedere leerling zijn of haar antwoorden
3. Tijdens het maken van het schoolwerk vraag ik de leerlingen hoe ze vinden dat het tot nu toe gaat
4. Ik laat mijn leerlingen meedenken over de manier waarop ze willen leren op school.
5. Ik geef leerlingen de gelegenheid zelf hun leerpunten te bepalen
6. Ik vraag mijn leerlingen om aan te geven wat ze goed en minder goed hebben gedaan in hun schoolwerk.
7. Ik stimuleer leerlingen om terug te kijken op hun leerproces en om te bedenken wat ze een volgende keer beter anders kunnen doen.
8. Ik laat mijn leerlingen weten wat hun sterke punten zijn op het gebied van leren.
9. Ik laat mijn leerlingen weten wat hun zwakke punten zijn op het gebied van leren.
10. Ik moedig mijn leerlingen aan om hun leerproces te verbeteren.
11. Ik geef mijn leerlingen aanwijzingen die hen helpen bij het leren.
12. Ik bespreek het gemaakte werk met mijn leerlingen zodat zij de lesstof beter begrijpen.
13. Ik bespreek met mijn leerlingen hun vorderingen.
14. Na een toetsmoment vertel ik mijn leerlingen hoe zij hun zwakke prestaties kunnen verbeteren.
15. Ik bespreek met mijn leerlingen hoe zij hun sterke kanten kunnen gebruiken om hun werk te verbeteren.
16. Ik stel samen met mijn leerlingen een strategie vast om hun zwakke punten te verbeteren

Scaffolding

1. Als ik merk dat leerlingen een onderdeel niet begrijpen pas ik mijn instructie aan.
2. Ik geef de leerlingen aanwijzingen om hen te helpen de lesstof te begrijpen.

3. Tijdens de les kunnen de leerlingen laten zien wat ze hebben geleerd
4. Ik stel de vraag op een begrijpelijke manier
5. Door de vragen te stellen tijdens de les help ik leerlingen om de lessof begrijpen.
6. Ik sta open voor de inbreng van mijn leerlingen in de klas.
7. Ik ga met mijn leerlingen in discussie over de antwoorden.
8. Ik zorg ervoor dat de leerling weet aan welke punten hij of zij moet werken om zijn of haar resultaten te verbeteren.
9. Ik geef leerlingen de mogelijkheid om vragen te stellen.
10. De leerling weet aan welke eisen zijn of haar werk moet voldoen.
11. Ik zorg ervoor dat de leerlingen weten wat ze kunnen leren van een opdracht.
12. Ik kan zien of de leerling het leerdoel heeft behaald door zijn werk

2.7 Appendix C. Items and subscales of the SAFL-Q (Dutch)

Monitoring

1. De leerkracht stimuleert mij om na te denken over hoe ik mijn schoolwerk kan verbeteren.
2. Nadat de leerkracht mijn toets heeft nagekeken bespreken we samen mijn antwoorden.
3. Tijdens het maken van mijn schoolwerk vraagt de leerkracht hoe ik vind dat het tot nu toe gaat.
4. De leerkracht laat mij meedenken over de manier waarop ik wil leren op school.
5. Ik krijg van de leerkracht de mogelijkheid om te bepalen wat mijn leerpunten zijn.
6. Mijn leerkracht vraagt mij wat ik goed en minder goed heb gedaan in mijn schoolwerk.
7. De leerkracht stimuleert mij om terug te kijken op mijn leerproces en om te bedenken wat ik een volgende keer anders kan doen.
8. De leerkracht geeft mijn sterke punten aan op het gebied van leren
9. De leerkracht geeft mijn zwakke punten aan op het gebied van leren.
10. Ik word aangemoedigd door mijn leerkracht om mijn leerproces te verbeteren.
11. Ik krijg aanwijzingen van de leerkracht die mij helpen bij het leren.
12. Mijn leerkracht bespreekt mijn gemaakte werk met mij zodat ik de lesstof beter begrijp.
13. Mijn leerkracht bespreekt met mij mijn vorderingen
14. Na een toetsmoment laat mijn leerkracht mij weten hoe ik het de volgende keer beter kan doen.
15. Mijn leerkracht bespreekt met mij hoe ik mijn sterke kanten kan gebruiken om mijn werk te verbeteren.
16. Samen met mijn leerkracht bedenk ik een manier om mijn zwakke punten te verbeteren.

Scaffolding

1. Als ik de uitleg niet begrijp dan probeert de leerkracht het op een andere manier aan mij uit te leggen.
2. De leerkracht geeft mij aanwijzingen die mij helpen om de lesstof te begrijpen.

3. Tijdens de les kan ik laten zien wat ze hebben geleerd
4. De leerkracht stelt de vraag op een begrijpelijke manier.
5. De vragen van de leerkracht helpen mij de lesstof te begrijpen
6. De leerkracht staat open voor mijn inbreng in de klas.
7. Ik heb de mogelijkheid om vragen te stellen aan medeleerlingen over de les.
8. Ik weet aan welke punten ik moet werken om mijn resultaten te verbeteren.
9. Er is mogelijkheid om vragen te stellen.
10. Ik weet aan welke eisen mijn werk moet voldoen.
11. Als ik een opdracht krijg is het duidelijk wat ik hiervan kan leren.
12. Met mijn werk laat ik zien wat ik kan.

3

Multilevel Predictors of Differing Perceptions of Assessment For Learning Practices Between Teachers and Students[2]

Abstract

Assessment for Learning (AfL), as a way to promote learning, requires a 'match', or shared focus between student and teacher to be effective. But students and teachers may differ in their perception of the purpose and process of classroom assessment meant to promote learning. Perceptions regarding AfL practices in their classroom were collected from both teachers and students. Teacher efficacy for instructional strategies, student engagement, classroom management, and student self-reported language proficiency were considered possible influencing factors. Multilevel analysis on self-report questionnaires on AfL practices administered to 650 students and 38 teachers revealed a substantial mismatch in perceptions between teachers and students. Congruency of teacher-student perceptions was highly homogeneous within classes. High teacher efficacy and low student language proficiency were associated with a incongruent AfL perceptions. Findings are interpreted using the self-verification theory.

3.1 Introduction

Incongruent perceptions of Assessment for Learning practice

Various authors (P. Black & Wiliam, 1998b) have advocated using assessment as an instructional approach to improve learning processes by utilizing its possibilities for scaffolding and monitoring student progress. This integration of assessments in the learning process has been called Assessment for Learning (AfL) and is contrasted against the traditional Assessment of Learning (P. Black & Wiliam, 1998a). Although Assessment for Learning may be part and parcel of classroom instruction; students and teachers may disagree on whether it is practiced (Broadfoot, 1998). A core principle of AfL is that it functions as a two-way learning process: learners are scaffolded into learning how to progress based on their current achievement, while teachers adapt their teaching according to the garnered assessment information of the current level. The

[2]This chapter is submitted for publication as: Pat-El, R.J., Tillema, H., Segers, M. & Vedder, P. (Under revision). Multilevel predictors of (mis)matching perceptions of Assessment for Learning practice: Teacher-efficacy and students' language proficiency.

extent to which AfL effectively supports students' learning is facilitated by a close 'match' or agreement between student and teacher about how assessment is utilized; students need to recognize teachers' efforts in scaffolding their learning and monitoring their progress in order to advance their learning (Popham, 2008; Sadler, 2010).

Some studies cast doubts on the congruency between teacher and student perception of AfL-practice (Birenbaum et al., 2006; McMillan, 2007), referring to conflicting perceived purposes (Popham, 2008) or stating that students preempt a formative use of assessment. Only scarce quantitative results are to be found in regards to the degree to which students and teachers align in their perception of practiced AfL. Clarifying possible mismatches between student-teacher perceptions could, in our view, contribute to the implementation of AfL. The aim of this study, therefore, is to evaluate student and teacher perceptions of actual AfL practices and find factors that may explain possible individual variance between students and teachers.

Focus on Assessment for Learning

Assessment can provide a powerful tool in enhancing learning and promoting students' motivation to learn (James & Pedder, 2006; McMillan, 2007). AfL is a process of continual interaction between teachers and individual learners, in which feedback provision and its acceptance and utilization are key elements (P. Black & Wiliam, 2009; Struyven et al., 2005; Davis, 2006). A necessary condition for students to accept feedback is to recognize that feedback is being provided. In defining the construct of AfL, common principles in most studies are identified as: (1) rich classroom questioning, (2) facilitating self- and peer assessments, (3) clarifying goals and criteria, and (4) giving feedback/feedforward (P. Black & Wiliam, 1998a; Assessment Reform Group, 2002; James & Pedder, 2006; Whitelock, 2010). As suggested by Stiggins (2005) and Pat-El, Segers, Tillema and Vedder (2011) these four general principles can be further subsumed under two functions of AfL: a) monitoring to track student progress and b) scaffolding to show or help students recognize what areas need improvement. Monitoring refers to analyzing student learning progress to foster student self-monitoring with the intent to find challenges and opportunities for optimizing learning. Scaffolding refers to classroom interaction wherein learning goals and criteria are clarified through, and in addition to, classroom questioning. These two ingredients constitute ways of feedback provision (Wiliam, 2011). In order for AfL to be supportive of learning, it needs to be recognized by students, and congruent perceptions between students and teachers are needed to optimize the benefits of feedback provision. Incongruent perceptions on whether Assessments are embedded in the learning process between teachers and students may lead to misunderstanding and misinterpretation of the assessment information (Bartholomew et al., 2001; Norman, 1986). Several studies (MacLellan, 2001; Raviv, Raviv, & Reisel, 1990) found differing perceptions of the formative nature of assessments between students and their teachers, in which teachers tend to perceive their assessments as formative, whereas students report more summative assessment methods. Students evaluate the formative nature of assessments to be more implicit and 'hidden' than their teachers (Könings, 2007). Moreover, students often reported they did not convincingly perceive clear goals, room for personalization of learning, or

fascinating subject-contents, teachers reported they did perceive these aspects to be present in daily educational praxis (Könings, 2007).

Unclear are the factors that influence the congruency of perceived AfL. The current study looks at promising explanatory constructs: (1) teachers' self reflective barriers through teacher efficacy, which refers to teachers' beliefs in their capability to successfully accomplish teaching tasks, (Allinder, 1994) and (2) students' language proficiency as a proxy for the students' understanding and appreciating AfL information.

Teacher Efficacy

Teachers' beliefs, particularly about their personal effectiveness or efficacy, have been found to contribute to their effectiveness and goal attainment (Ashton & Webb, 1986). Teacher with high efficacy were oriented towards high students' achievement and success (Guskey, 1988). Perceived self-efficacy has been defined as personal judgments of one's competency to produce desired effects by one's actions (Bandura, 1997). Tschannen-Morann and Hoy (1998) defined three domains of teacher-specific competency beliefs: (1) efficacy for instructional strategies, (i.e., teachers' belief in being able to use varied assessment strategies, respond to difficult questions, or engage students at adequate levels of competence); (2) efficacy for classroom management, (i.e., maintaining classroom rules and order); and (3) efficacy for student engagement, (i.e., getting students motivated to learn). Teacher efficacy has been linked to effective classroom instruction (Eren, 2009), openness to new ideas (Allinder, 1994), and enthusiasm for classroom instruction (Hall, Burley, Villeme, & Brockmeijer, 1992). However, Gerges (2001) showed that teacher efficacy may block flexibility and variation in employing instructional strategies, such as explorative classroom questioning. High efficacious teachers seem to show more rigid use of teaching strategies and content coverage, (Wheatly, 2002). Mastery experiences contribute to beliefs of efficacy (Bandura, 1997), but mastery experiences themselves remain a subjective experience, wherein self-efficacious beliefs themselves can fuel a confirmation bias for mastery experiences, in turn boosting one's own self-efficacious beliefs (Mynatt, Doherty, & Tweney, 1977). This process of self-verification (Swann Jr., Chang-Schneider, & Angulo, 2007) describes how people are more apt to seek information that confirms one's self-beliefs, in order to avoid incongruency between their perceived self and their experienced self (Carver & Scheier, 2000). In this respect, teachers high on teacher efficacy beliefs would focus more on particular classroom signals that boost their efficacy beliefs, and focus less on information that is contrary to those. The prime focus on congruency with personal intentions might run the risk of missing student signals important for alignment in assessment perceptions. AfL requires teachers to actively adapt their teaching based on student information. Thus, teachers high on teacher efficacy beliefs might be at risk of missing out on cues signaling them to adapt their teaching. In line with the theory of self-verification we expect teachers with high teacher efficacy beliefs, mainly beliefs concerned with the successful implementation instructional strategies, to perceive more AfL practiced in their classroom than their students do.

Language proficiency mediating assessment for learning

Formative feedback, as key ingredient of AfL, needs to be understood by the student and recognized as supportive for learning (Bartholomew et al., 2001). Assessment practices in classrooms build on the students' ability to grasp meaning and purpose of the information provided to them by, in this case, the AfL strategies of monitoring (i.e., appreciation of strength and weaknesses) and scaffolding (action on learning routes) (Sadler, 2010). We assume that student language proficiency becomes an important determinant to understanding assessment information (Wertsch, 1997). Failing to grasp any nuances in feedback provision can become detrimental when students misunderstand their teacher's communication.

Student language proficiency has been found to be strongly related to learning success (Oortwijn, Boekaerts, & Vedder, 2008), to student approaches to learning (Biggs, 1990), and academic achievement (Collier, 1989; Van der Slik, Driessen, & De Bot, 2006). AfL, with its provision of formative feedback, is a linguistic endeavor. The importance of language proficiency in an academic context has been stressed as to apply to all students regardless of their cultural backgrounds. In the current study, we take up the important role of language for learning, to hypothesize that students, low on language proficiency will less appreciate AfL information offered to them. .

The current study

The current study investigates the alignment of students' and teachers' perceptions on AfL as practiced within their classroom. Our research question is: do students and teachers differ in their perceptions of AfL practice, and if so, can those differences be explained by high teacher efficacy and low student language proficiency?

3.2 Method

Sample

Thirty-one junior vocational high schools in the Netherlands were randomly approached, of which seven agreed to participate in this study. Students and teachers of these schools took part in this study. Questionnaires in this study were administered to 650 students (N_{girls} = 323, N_{boys} = 326; 1 missing) and 38 teachers. The participating teachers represented a broad domain of subjects. The average class size was 17.6 students (SD = 4.88; min = 11; max = 27). Teachers (20 females, 18 males) were on average 44.5 years old (SD = 11.21). and active for 17.8 years (SD = 12.37) (at their current school for 10.0 years (SD = 10.32). Female teachers (M = 42.5, SD = 9.68) and male teachers (M = 46.6, SD = 12.62) were about the same age, ($t(36)$ = -1.13, p = .264). Students were on average 13.9 years old (SD = 1.13); girls (M = 13.7, SD = 1.10) were slightly younger than boys (M = 14.1, SD = 1.13), $t(643)$ = -4.17, $p < .001$.

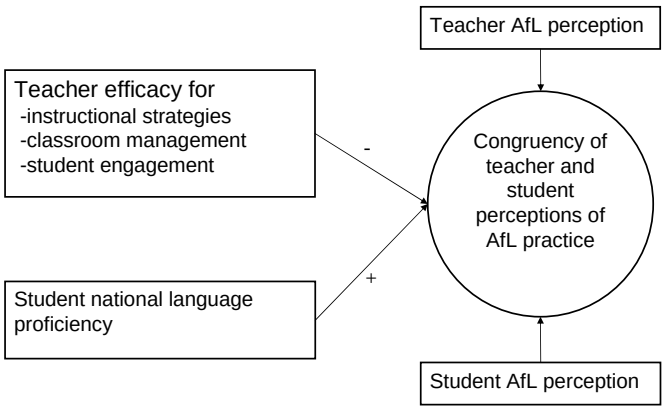


Figure 3.1: Theoretical model

Measures

Questionnaires were used to measure AfL perceptions in teachers and students in conjunction with teachers' efficacy for teaching, teachers' and students' self-esteem, students' language proficiency, as well as demographic variables such as age, and gender.

Dependent variable

Perception of AfL.

Perceptions of AfL practices were measured with the Assessment for Learning Questionnaires which consists of both a teacher and a student version (Pat-El, Tillema, et al., 2011). This questionnaire was validated to compare teacher and student perceptions of classroom AfL practice. The questionnaire consists of two subscales: Monitoring, which is comprised of items about feedback giving, and Scaffolding, which is comprised of items about clarifying learning goals and criteria. All items were scored on 5-point Likert scale items. Differences (i.e., alignment) were computed by subtracting the teachers' TAFLQ mean subscale scores from the student scores from the SAFL-Q. Negative difference scores therefore indicate higher teacher scores, while positive difference scores indicate higher student scores. Items in both the TAFL-Q and the SAFL-Q are worded similarly.The Monitoring subscale consisted of 12 items (Cronbach's α = .91). Sample items are "I encourage my students to reflect upon how they can improve their assignments", and "I discuss with my students how to utilize

their strengths to improve on their assignments". The Scaffolding subscale consisted of 16 items (Cronbach's $\alpha = .87$). Sample items are "I ensure that my students know what they can learn from their assignments", and "I adjust my instruction whenever I notice that my students do not understand a topic".

Independent variables

Teacher efficacy.

Teachers rated their efficacy for teaching on the Ohio State Teacher Efficacy Scale (OSTES) (Tschannen-Morann & Hoy, 1998). The 12-item OSTES measures teacher efficacy in three subscales: Efficacy for instructional strategies (EIS), efficacy for classroom management (ECM), and efficacy for student engagement (ESE). Each subscale was comprised of four self-report Likert items. To avoid different measurement scales in the questionnaire, the measures were adapted to five-point scales ranging from 1 (strongly disagree) to 5 (strongly agree). Sample items are "To what extent can you provide an alternative explanation or example when students are confused" (EIS); "How much can you do to control disruptive behavior in the classroom" (ECM); and "How much can you do to get students to believe they can do well in schoolwork" (ESE). Cronbach's αs in the present study were .77, .88 and .73 respectively. α for the total self-efficacy score was .85.

Language proficiency.

Student language proficiency was measured with a self-report scale from the ICSEY study (Berry, Phinney, Sam, & Vedder, 2006). This four-item self-report scale to determine language proficiency showed high correlations with evaluations of a person's language proficiency by others (Kirchmeyer, 1993). In the four-item scale students were asked to evaluate how well they were able to read, write, speak and understand the Dutch language. Scores ranged from 1 (not at all) to 5 (very well). Cronbach's α in the present study was .82.

Control measurement

Since male and female students and teachers generally differ in their evaluation of the learning environment (Van Grinsven & Tillema, 2006) this study controlled for student and teacher gender. Since Self-esteem of teachers and students is strongly related to teacher efficacy beliefs (Judge & Bono, 2001), and with students' evaluations of feedback (Nicol & MacFarlane-Dick, 2006), student and teacher self-esteem will be used as control variables as well.

Students and teachers rated their self-esteem with the Single Item Self-Esteem measure (Robins, Hendin, & Trzesniewski, 2001), The item is "I have high self-esteem" scored on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Procedure

Schools were approached by email and telephone. Teachers and their students participated by informed consent. Students completed the questionnaires during regular school hours. Filling out the questionnaires took teachers and stu-

dents about 25 minutes. Respondents were assured that their contribution was anonymous. Students received a candy, as small reward when they returned their fully completed questionnaire.

Analysis

The design used was cross-sectional and correlational. Multilevel modelling (in MLwin version 2.02) was used to deal with the hierarchical nature of the data. In this way classroom effects can be controlled for when analyzing student effects, and classroom level variables can be introduced to the models (Kreft & De Leeuw, 1998). Differences in perceptions between students and teachers were first calculated into a difference score. This difference score was used as an outcome variable in a multilevel regression model. In the null-model, where only the constant was used in the equation, this procedure allows for the testing of mean differences between teachers and their specific students. Multilevel backwards selection was used to test both predictors and interaction between predictors in explaining the AfL perception alignment between teachers and students.

3.3 Results

A descriptive summary of the variables used in this study is provided in Tables 3.1 and 3.2. Student and teacher means and standard deviation on the perception of AfL practice indicated that teacher means are substantially higher than student means, with little variation within groups. Teachers evaluated themselves as highly self-efficacious.

Do teacher and student perceptions of AfL differ?

Multilevel modeling was used to determine matching in perceptions of AfL. The null-model, where only the constant was entered in the model, was used to evaluate the mean of perception differences. Teachers (M = 3.8; SD = 0.61) reported more use of monitoring than their students (M = 2.8; SD = 0.76), β = -0.93 (0.11), $p < .001$) and more use of scaffolding (M = 4.2; SD = 0.49) than their students (M = 3.5; SD = 0.81), β = -0.70 (0.11), $p <$ .001). The intra-class correlation for differences in Monitoring was .51 and for differences in Scaffolding .48. The high intra-class correlations showed that average differences between students and teachers can vary greatly from teacher to teacher, and that alignment in perceptions is highly homogeneous within classrooms. This indicates that there is a strong relationship between AfL perceptions in a classroom, instead of it being a mostly individual mismatch. There was no indication of a random slopes model.

Predicting AfL perception differences

Predictors of AfL perception differences were tested with a multilevel backwards selection method (Twisk, 2006). The starting model has all variables and all necessary random intercepts and slopes included. In each subsequent step the predictor with the lowest p-value is deleted until only predictors with

Table 3.1: Means and standard deviations of the variables in the present study.

Variable	N	Mean	SD
Student level variables			
Student age	650	13.9	1.13
Monitoring	650	2.8	0.76
Scaffolding	650	3.5	0.81
Student self-esteem	643	3.5	1.17
Dutch Language proficiency	643	4.5	0.65
Teacher level variables			
Monitoring	38	3.8	0.61
Scaffolding	38	4.2	0.49
Teacher self-esteem	37	4.1	0.85
Years teaching	38	17.8	11.65
Years teaching in current school	37	9.8	9.71
Teacher age	38	44.5	11.21
Efficacy for instructional strategies	38	4.2	0.54
Efficacy for classroom management	38	4.2	0.70
Efficacy for student engagement	38	3.6	0.63
Perception differences			
Monitoring Student - Teacher	648	-1.0	.89
Scaffolding Student - Teacher	638	-0.7	.96

$p < .10$ remain. In Table 3.3 and Table 3.4 a summary of the starting model and the final model are given.

Monitoring.

After first entering control variables and predictors, backwards selections yielded the following results: low degree of alignment (large student- teacher difference on AfL per classroom) on Monitoring was significantly predicted by Student Language Proficiency, Efficacy for Instructional Strategies (EIS) and teacher gender. Only student self-esteem and student gender showed significant changes in deviance (Δ) when random slopes were removed, indicating that the average classroom-level differences of these two variables should remain in the model. (Table 3.3)

The final model indicates that higher Student language proficiency (β = 0.07, p = .04) corresponds to a slightly closer alignment between teachers' and students' perceptions of Monitoring, and higher teacher efficacy in instructional strategies corresponds to a wider gap (β = -0.43, p = .005). The relationship between EIS and monitoring alignment was inspected for curvilinearity, but the relationship was found to be linear. The apparent mismatch was found to be larger for female teachers than for males (β = -0.41, p = .006). Adding the interaction between teacher gender and efficacy for instructional strategies did not yield a significant improvement in fit (β = -0.15, p = .31, Δdeviance = -.29, p = .65). Efficacy for student engagement was a marginally significant predictor for mismatches in monitoring perceptions. Efficacy for classroom

Table 3.2: Correlation matrix of the variables in the present study.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Student age	1.00														
2	Student-Monitoring	0.14	1.00													
3	Student-Scaffolding	0.12	0.70	1.00												
4	Student self-esteem	-0.02	0.10	0.13	1.00											
5	Dutch Language proficiency	-0.02	0.07	0.14	-0.06	1.00										
6	Teacher-Monitor	0.24	0.16	0.05	-0.02	-0.06	1.00									
7	Teacher-Scaffold	0.23	0.04	-0.02	-0.01	-0.11	0.73	1.00								
8	Teacher self-esteem	-0.09	-0.02	0.03	0.00	0.00	0.10	0.17	1.00							
9	Years teaching	0.39	0.01	-0.05	0.06	-0.02	0.22	0.11	0.08	1.00						
10	Years teaching in current school	0.29	0.00	0.00	0.04	0.05	0.24	0.20	0.04	0.58	1.00					
11	Teacher age	0.13	0.01	-0.03	0.07	0.01	0.01	-0.08	0.31	0.85	0.61	1.00				
12	Efficacy for instructional strategies	0.17	0.17	0.11	0.00	-0.05	0.62	0.53	0.18	0.03	0.03	-0.06	1.00			
13	Efficacy for classroom management	0.25	0.14	0.16	-0.01	-0.12	0.46	0.49	0.43	0.11	0.06	0.04	0.40	1.00		
14	Efficacy for student engagement	0.21	0.12	0.08	-0.05	-0.11	0.60	0.58	0.10	0.03	0.23	-0.08	0.58	0.37	1.00	
15	Monitoring Student - Teacher	-0.04	0.75	0.56	0.09	0.10	-0.53	-0.46	-0.13	-0.13	-0.16	0.01	-0.27	-0.19	-0.30	1.00
16	Scaffolding Student - Teacher	-0.02	0.56	0.85	0.11	0.17	-0.35	-0.55	-0.03	-0.10	-0.11	0.02	-0.20	-0.13	-0.24	0.72

management (ECM) was not found to be a significant predictor of differing perceptions on Monitoring .

Scaffolding.

After first entering all variables, both control and predictors, backwards selections yielded the following results: A high degree of difference was significantly predicted by Student self-esteem, Language proficiency, and Teacher Efficacy for instructional strategies (Table 3.4). Only student self-esteem and student gender showed significant changes in deviance when random slopes were removed.

The results indicate that high language proficiency scores correspond with a smaller difference between teachers' and students' perceptions of scaffolding (β = 0.13, $p < .0001$), while on the other hand teachers' high in efficacy in instructional strategies correspond with larger discrepancy scores (β = -0.438, $p < .0001$). The relationship between EIS and scaffolding alignment was inspected for curvilinearity, but the relationship was found to be linear. High student self-esteem coincided with a slight reduction in discrepancy (i.e., reducing mismatch) (β = 0.07, $p = .006$). Efficacy for student engagement and efficacy for classroom management were not significant predictors of difference in perception on Scaffolding.

Table 3.3: Backwards selection of predictors for the difference scores of Monitoring (student-teacher).

	Null β(SE)	Start β(SE)	final β(SE)
Intercept	-0.93^a*(0.11)	1.34^b*(0.73)	1.47^b*(0.69)
Student Level			
Student gender (boy = 0)		0.08(0.09)	
Interaction Gender student*teacher		-0.14(0.13)	
Student self-esteem*		0.04(0.03)	0.05^c*(0.03)
Dutch Language proficiency		0.06(0.04)	0.07^b(0.04)
Teacher level			
Teacher gender (male = 0)		-0.42(0.18)	-0.41^a(0.16)
Teacher self-esteem		0.01(0.11)	
Efficacy for instructional strategies		-0.50(0.20)	-0.43^b(0.19)
Efficacy for classroom management		0.09(0.15)	
Efficacy for student engagement		-0.22(0.17)	-0.25^c(0.16)
		u1	0.15
		u2	<0.01 0.01
		e	0.36
deviance	1295.93	1190.96^a	1247.18^a

Note: $^a = p < .01$; $^b = p < .05$; $^c = p < .10$; * = random intercept

3.4 Discussion

The aim of this study was to examine the congruency of student and teacher perceptions of AfL practices and to examine whether possible teacher-student

Table 3.4: Backwards selection of predictors for the difference scores of Scaffolding (student-teacher).

	Null β(SE)	Start β(SE)	final βSE
Intercept*	-0.70[a]*(0.11)	0.26*(0.89)	0.36*(0.78)
Student Level			
Student gender (boy = 0)*		-0.13(0.08)	-0.13[c]*(0.08)
Interaction Gender student*teacher		0.01(0.12)	
Student self-esteem*		0.07(0.02)	0.07[a]*(0.03)
Dutch Language proficiency		0.14(0.04)	0.13[a](0.04)
Teacher level			
Teacher gender (male = 0)		-0.37(0.23)	
Teacher self-esteem		0.06(0.14)	
Efficacy for instructional strategies		-0.27(0.24)	-0.44[b](0.18)
Efficacy for classroom management		0.06(0.18)	
Efficacy for student engagement		-0.26(0.20)	
	u1	0.18	0.20
	u2	0.02 0.11	0.03 0.10
	u3	<0.01 -0.01 0.01	0.01 -0.01 0.01
	e	0.39	0.36
deviance	1395.70	1261.57[a]	1316.85[b]

Note: [a] = $p < .01$; [b] = $p < .05$; [c] = $p < .10$; * = random intercept

differences could be attributed to teacher efficacy and student language proficiency. Mismatching perceptions between students and teachers and its predictors have been rarely quantitatively researched but are of importance to understand the impact Assessment for Learning has on promoting student learning (James & Pedder, 2006; McMillan, 2007). Based on the theory of self-verification (Swann Jr. et al., 2007) it was expected that efficacious teachers focus on information that confirm their self-efficacious beliefs regarding instructional and management routes taken in instruction, which would predict a high end perception of their AfL practice in comparison to what their students perceive. On the student level, language proficiency was seen as a predicting factor in appreciating and understanding the supportive function of feedback.

This study's findings confirmed differences in student and teacher perceptions of AfL-practice within classrooms. Teachers in general perceived a higher level of AfL practice as present in their classrooms than their students did. The incongruencies found on Monitoring and Scaffolding varied greatly between classrooms and is highly homogeneous within classrooms, which suggests that there is a strong relation between a teacher's teaching and AfL practice on the one hand and student perception of it on the other. It is therefore likely that teaching-related variables could explain disagreements between students and teachers. The mismatch found in our study supports findings in earlier studies (Könings, 2007; MacLellan, 2001) in which teachers were found to overestimate their practice of AfL. Further study into mismatching perceptions on the practice of AfL between teachers and their students can help to detect why some instructional activities to foster learning sometimes have unexpected or contradictory effects (Olkinuora & Salonen, 1992).

In our study, teacher efficacy for instructional strategies and students' language proficiency predicted differences in AfL perceptions: high teacher efficacy was associated with more incongruency in AfL. In line with the theory of self-verification (Swann Jr. et al., 2007), efficacious beliefs about instructional design, and to a lesser degree about student engagement, predict a larger incongruency between teachers' and students' perception of AfL as it is practiced. The relationship we found between efficacy and AfL supports the argument that teachers who are confident in their ability to teach might focus more on the feedback they give, and subsequently may overestimate their students' evaluations of AfL.

Student language proficiency also was found to predict AfL perception differences. That is: students' higher levels of self-reported language proficiency correspond to more congruency. The relationship between language proficiency and AfL incongruency was stronger for Scaffolding than for Monitoring. These findings suggest that scaffolding is more language dependent than monitoring. For male students the mismatch was marginally bigger than for female students. An explanation may be that Scaffolding in AfL depends on understanding and following recommendations from feedback. Studies are available in which boys tend to appraise teaching and processes of evaluations more negatively (Worthington, 2002), and that girls are more language proficient than boys (Boyle, 1987). A more substantive explanation is that high instructional efficacious (i.e., task oriented) teachers initiate more negative interactions with male students than with female students (Jones & Dindia, 2004).

We found gender related differences also to be present for the teachers participating in this study. A greater mismatch in perceived monitoring was found especially for female teachers. The analysis suggests that this effect is not due to an interaction of teacher gender with efficacy for instructional strategies. A meta-analysis by Cornelius-White (2007) might help explain this finding. This meta-analysis suggests that female teachers are more likely than male teachers to emphasize the social-emotional aspects of classroom instruction (and assessment for that matter). Although it remains conjecture, because scaffolding is more associated with social-emotional aspects of teaching, through classroom dialogue and sharing understanding of personal goals to realize growth (Järvelä, Lehtinen, & Salonen, 2000), it could be that female teachers are more able to correctly recognize socio-emotional signals, but have more difficulties in task-related areas corresponding with monitoring activities.

The exploratory nature of the study was reflected by its cross-sectional correlational design. Theoretically we have support for the hypothesis that teacher efficacy, through a process of self-verification, can be an explanation for misalignments between students and teachers. This study's main findings have implications for classroom practices and research in AfL. Teacher efficacy seems a double edged sword; it can be a positive factor in valuing a mastery goal structure (Wolters & Daugherty, 2007), but also work out negatively by blocking variation and by directing teachers' focus away from (implicit) student learning signals.

4

Teacher-Student perceptions of assessment practices as predictors of student motivation to learn[3]

Abstract

Assessment in classrooms is a highly promotional tool for learning but often feared for its summative nature. The aim of this study was to test the hypothesis that differences between students and teachers in the perception of monitoring and scaffolding activities would predict students' intrinsic motivation as mediated by the students' basic needs of Competence, Autonomy and Relatedness. It was expected that students who are more proficient in the language of instruction would be more congruent with their teachers, and that teachers who are more efficacious in their teaching would be less congruent with their students. A total of 1466 students and 89 teachers from junior vocational high schools participated in this survey-study. Multilevel structural equation modelling revealed that differences between teachers' and students' classroom perceptions of AfL were smaller with efficacious teachers, and positively predicted intrinsic motivation which was partly mediated through basic motivational need fulfilment.

4.1 Introduction

Fostering student motivation through assessment and instruction is generally considered important to achieve positive learning outcomes (see Harlen & Crick, 2003, for a review). Student motivation positively predicts learning achievements (Boekaerts & Corno, 2005; Linnenbrink & Pintrich, 2002) and has been found to correlate with student competency beliefs (Ames & Archer, 1988), perceived school well-being (Kasser & Ryan, 2001), and student dropout (Legault, Green-Demers, & Pelletier, 2006). Given the body of research on fostering student motivation, it is surprising that the impact of learning scaffolding tools such as assessment for learning on student motivation has not yet been studied extensively by educational researchers.

[3]This chapter is submitted for publication as: Pat-El, R. J., Tillema, H. H., Segers, M.S.R. & Vedder, P. H. (Under Review). Teacher-student perceptions of assessment practices as predictors of student motivation to learn.

Teacher and student perceptions of Assessment for Learning

The repertoire of formative, learning enhancing assessment methods has increased the past decades into a diversified field of alternative assessment tools, such as teacher-initiated- , self- and peer assessments using a variety of assessment techniques such as closed questions, essays, portfolios and performance assessment in authentic learning situations. Assessment for Learning (AfL), as an approach, stresses the importance of continuous monitoring and repeated provision of informative feedback and classroom dialogue to promote student learning (James & Pedder, 2006). Various authors (P. Black & Wiliam, 1998b) have advocated that using assessment as an instructional approach will promote (deep) learning processes by utilizing its possibilities for monitoring to track student progress and scaffolding to show or help students recognize what areas need improvement (Pat-El, Tillema, et al., 2011; Stiggins, 2005). Advocates of the AfL approach point to the need for integrating assessment within learning and thereby putting emphasis on formative assessment as supportive to the learning process (McMillan, 2007). If AfL is to integrate assessment with learning, it has to be a process of continual interaction between teachers and individual learners in which feedback provision and its acceptance and utilization are key elements (P. Black & Wiliam, 2009; Davis, 2006; Struyven et al., 2005). In this respect, it is of great importance that students' and teachers' perceptions on the nature and content of the assessment provided are congruent. In order for teacher assessment to feed in to student learning, the teacher may have to adapt word choice and complexity of the information entailed in the feedback to the student's capacities to understand the feedback. This means that the teacher seeks to enrich the students' learning without disturbing student's mood and attentional focus for learning (P. H. Vedder, 1985). Obviously, this cannot be a one-sided, teacher regulated process. The student needs to disclose how she learns, what the contents are that she is working with, and how she understands task or assignments that are supposed to guide the learning process. Also, it is important that the student anticipates teacher instructions and feedback as personal expectations. Student expectations of the teacher's contribution prepare the student for more or less optimal 'absorption' or inclusion of the teacher provided information into the flow of learning. The less congruent teachers and students are in what each of these parties has to know about the information needed to optimize instruction and learning, the less effective instruction and learning will be. Congruency in teacher and student assessment and learning related perceptions are important for the success of classroom interventional activities (Loughran, 2010). Students who perceive teachers as having failed to provide support show less interest and enjoyment in school (Skinner & Belmont, 1993). Research shows that assessment practices can easily distract students from learning (Doyle, 1977), particularly when assessment practices affect students' feelings of safety in their setting of learning (Covington, 2000; Entwistle & Tait, 1990). Misaligned perceptions of the learning environment lead to misunderstanding and possible misinterpretations of the assessment information; its meaning and purpose (Norman, 1986; Bartholomew et al., 2001), and students can perceive the learning criteria to be more implicit and 'hidden' than their teachers (Könings, 2007).

Recent research on teacher or student perceptions of their learning environment, shows that there is a clear misalignment between teachers and students

on the focus and direction of (in)formative feedback support (Hattie, 2008; MacLellan, 2001; McMillan, 2007; Pat-El, Tillema, Segers, & Vedder, 2010). The fact that research showed that teacher-reported teaching behavior can strongly differ from actual behavior (Reeve, Bolt, & Cai, 1999) suggests that it may be complicated to correct or compensate for misalignments.

Predicting perception congruencies

Creating a learning environment that is facilitative to students' learning and motivation requires considerable teaching effort. This teacher effort depends on experiences of teacher efficacy or teacher's sense of being effective, a good teacher (Allinder, 1994). Teacher efficacy might therefore prove to be an important predictor of congruence of student and teacher appreciations of AfL components.

Indeed it may be an important predictor, but at the same time teacher efficacy is a potential fallacy. Teachers who are confident in their ability to teach might focus more on classroom feedback that supports their efficacious beliefs, and subsequently may overestimate their own perceptions of AfL. For example, Gerges (2001) showed that teacher efficacy may block flexibility and variation in employing instructional strategies, such as explorative classroom questioning. High efficacious teachers seem to show more rigid use of teaching strategies and content coverage (Wheatly, 2002). Self-efficacious beliefs themselves can fuel a confirmation bias for mastery experiences, in turn boosting one's own self-efficacious beliefs (Mynatt, Doherty, & Tweney, 1977). Through this process of self-verification (Swann Jr. et al., 2007) teachers high on teacher efficacy beliefs would focus more on particular classroom signals that boost their efficacy beliefs, and focus less on information that is contrary to those and risk missing out on cues signaling them to adapt their teaching. In line with the theory of self-verification we expect teachers with high teacher efficacy beliefs for successfully implementing instructional strategies (Tschannen-Morann & Hoy, 1998) to perceive more AfL practiced in their classroom than their students do. From a students' perspective, assessment practices in classrooms ideally build on the students' ability to grasp meaning and purpose of the information provided by the AfL strategies of monitoring (i.e., appreciation of strength and weaknesses) and scaffolding (action on learning routes) (Sadler, 2010).

In order to appreciate and understand assessment information and feedback given, student language proficiency is an important characteristic in classroom instruction (Wertsch, 1997). Formative feedback needs to be understood and recognized as support for it to be effective (Bartholomew et al., 2001). Feedback might not be recognized or understood and then it might not be experienced or valued as teacher support. Any nuances between plainly criticizing and feedback can become blurred when teachers and students misunderstand each other's communication possibly resulting in incongruent teacher-student perceptions of AfL.

The impact of assessment on motivation

Self Determination Theory is used in this study to explain motivational states as outcomes of differences between student and teacher perceptions of classroom assessment interventions. SDT proposes that students have to satiate

innate universal 'needs' to feel motivated for an educational activity (Deci & Ryan, 1985; Ryan & Deci, 2000). Intrinsic motivation is the tendency to engage in activities for the inherent joy an activity brings, and has been positively linked with persistence, mastery learning goals, deep learning, and well being (Ryan & Deci, 2000). Motivational needs consist, according to SDT, of a feeling of being autonomous, a sense of relatedness with others in the activity, and experiencing the competence to complete a given activity. These three needs are characterized by Jang (2008) as basic needs. This scholar presents some evidence that the basic needs in part explain the effectiveness of AfL in fostering student motivation. Scaffolding activities as provided in Assessment for Learning, especially giving direction and clarification of learning goals may successfully fulfil all three basic needs (Jang, 2008). Studies that have focused on certain motivational needs in the SDT, in the context of monitoring activities, have found that feedback is positively associated with intrinsic student motivation (A. E. Black & Deci, 2000). The provision of feedback and the support of student autonomy has also been related to students' increased feelings of competence and autonomy (Levesque, Zuehlke, Stanek, & Ryan, 2004), indirectly and positively affecting intrinsic student motivation (Jang, Reeve, Ryan, & Kim, 2009).

There is some evidence that teacher-student misalignments have an impact on motivational mediators. For example, instances where a high degree of agreement between teachers and students was found, students also reported a different, closer relationship with their teachers, and their role was more categorized as a 'teacher' than as a 'judge' (Sambell & McDowell, 1998). Students and teachers have also been found to disagree on the degree of autonomy support provided, or the importance attached to it (Assor et al., 2002), which negatively affected students' perceived personal preference for schoolwork. Feedback meant to foster student ability and feelings of competence have also been found to have the opposite effect when low ability students perceive the feedback as a teacher's doubt about their competence (P. Black & Wiliam, 1998a). Apparently assessment information can be 'lost in translation' in which teachers and students ultimately differently perceive the learning environment, which negatively impacts student motivation.

Because few studies using the SDT model of motivation incorporate all three basic needs it remains unclear in what way the effect of classroom assessment on intrinsic motivation is mediated by the basic needs. The aim of this study is to use the SDT model of intrinsic motivation and to investigate how congruency in the perception of scaffolding and monitoring predict intrinsic motivation, and whether that effect is mediated by basic need fulfillment.

The current study

The present study aims to test the congruency in the perception between teachers and students of the nature of the formative assessment provided to learners and whether the level of conguency positively corresponds with students' intrinsic motivation. Using self-determination theory of motivation as a conceptual background, it will be tested whether fulfillment of the basic needs of competence, autonomy and relatedness will explain the nature of the relationship between congruency in perception of AfL and intrinsic motivation. Teacher

efficacy and student language proficiency will be used as predictors of student-teacher congruent perceptions of the learning environment.[1]

4.2 Method

Sample

Students and teachers in junior vocational high schools in the Netherlands took part in this study. The individual students and their teachers were the unit of analysis. Questionnaires in this study were administered in eighteen schools to 1658 students (N_{girls} = 751, N_{boys} = 861; missing = 86) and 89 teachers. The participating teachers represented a broad domain of subjects, ranging from arts to sciences. The average class size was 17.6 students (SD = 4.88; min = 11; max = 27). Median students' age was 14 years old (min = 11; max = 19); girls (M = 13.7, SD= 1.10) were slightly younger than boys (M = 14.1, SD = 1.13), $t(643)$ = -4.17, $p < .001$. Teachers (43 females, 46 males) were on average 41.4 years old (SD = 11.97) and had been active for 14.7 years (SD = 11.81) and at their current school for 7.9 years (SD = 11.81). Female teachers (M = 39.4, SD = 10.80) and male teachers (M = 43.3, SD = 12.80) were about the same age, ($t(85)$ = -1.55, p = .126), but male teachers had more years of teaching experience (M_{female} = 11.9; M_{male} = 17.3; $t(85.6)$ = -2.23, p = .028) and more years of experience at their current school (M_{female} = 5.6; M_{male} = 10.2; $t(72.8)$ = -2.53, p = .014). Two teachers were removed from the analysis because their classes did not complete any questionnaires.

Measures

Questionnaires were used to measure AfL perceptions in teachers and students in conjunction with teachers' efficacy for teaching, students' language proficiency, students' three basic needs for motivation, and their intrinsic motivation. All Likert items response values ranged from 1 (strongly disagree) to 5 (strongly agree), unless otherwise specified.

Intrinsic motivation.

Intrinsic motivation was measured with the interest/enjoyment scale from the Intrinsic Motivation Inventory (IMI) (McAuley, Duncan, & Tammen, 1989). The scale was translated to Dutch and adapted to measure interest and enjoyment in the class they were being taught at that moment. The scale consisted of 7 Likert scale items. Sample items are: "I would describe this class as very interesting" and "I think this is a boring class" (reversed). In the current study, Cronbach's α of the 7-item scale was .91.

Perceived competence.

The 4-item Perceived Competence Scale (Deci, Schwartz, Sheinman, & Ryan, 1981) was translated to Dutch. A sample item is: "I am capable of learning

[1] Although teacher gender was a significant predictor of congruency in Chapter 3, due to lack of a clear explanatory theory this variable was not further tested. Non-significant predictors in Chapter 3, that were retained in the model to prevent overfitting, were not included in this and subsequent chapters

the material in this class". In the current study, Cronbach's α of the 4-item scale was .86.

Relatedness.

Relatedness was measured with the 8-item Relatedness scale from the IMI (Deci et al., 1981) and was translated to Dutch and adapted to measure the class they were being taught at that moment. A sample item is: "I feel like I can really trust my teacher". In the current study, Cronbach's α of the 4-item scale was .85.

Perceived autonomy.

The 3-item Perceived Autonomy Scale (Martens & Kirschner, 2004) was in Dutch. A sample item is: "I can determine for myself how I work during this class". In the current study, Cronbach's α of the 3-item scale was .67.

Independent variables

Perception of AfL.

Perceptions of AfL practices were measured with the Assessment for Learning Questionnaires which consists of both a teacher and a student version (Pat-El, Tillema, et al., 2011). The questionnaire is comprised of two subscales: Monitoring and Scaffolding. All items were scored on 5-point Likert scale items. Differences (i.e., alignment) were computed by subtracting the teacher scores, from the TAFL-Q, from the student scores, from the SAFL-Q. Negative difference scores therefore indicate higher teacher scores, while positive difference scores indicate higher student scores. Items in both the TAFL-Q and the SAFL-Q are worded similarly to enable strict comparisons between the two populations.The Monitoring subscale consisted of 12 items (Cronbach's α = .91). The scale was defined as how often and in what form feedback is used and how self-monitoring is facilitated. Sample items are "I encourage my students to reflect upon how they can improve their assignments", and "I discuss with my students how to utilize their strengths to improve on their assignment".

The Scaffolding subscale consisted of 16 items (Cronbach's α = .87). The scale was defined as communication of clear learning goals and criteria, and how those goals and criteria can be met. Sample items are "I ensure that my students know what they can learn from their assignments", and "I adjust my instruction whenever I notice that my students do not understand a topic".

Teacher efficacy for instructional strategies.

Teachers rated their efficacy for teaching on the Ohio State Teacher Efficacy Scale (OSTES) (Tschannen-Morann & Hoy, 1998). The 4-item Efficacy for instructional strategies was translated into Dutch. A sample item is, "To what extent can you provide an alternative explanation or example when students are confused" Cronbach's αs in the present study was .77.

Dutch language proficiency.

Student language proficiency was measured with a self-report scale from the ICSEY study (Berry et al., 2006). Self-reports to determine language proficiency have a high correlation with evaluations of a person's language proficiency by others (Kirchmeyer, 1993). In the four-item scale students were asked to evaluate how well they were able to read, write, speak and understand the Dutch language. Scores ranged from 1 (not at all) to 5 (very well). Cronbach's α in the present study was .82.

Procedure

Schools were randomly selected across the Netherlands by email and telephone. Of the 31 schools that were approached, 7 agreed to participate in this study. Teachers and their students participated by informed consent. Filling out the questionnaires took teachers and students about 25 minutes. Respondents were assured that their contribution was anonymous. Students received a small reward when they returned their fully completed questionnaire.

Analysis

Complex mediations are traditionally tested in structural equation modelling, but are generally inappropriate if the data is hierarchical in nature. The advantages of multilevel regression and structural equation modelling can be combined in a multilevel structural equation model (MSEM) (Muthén, 1994). In contrast to other methods of testing for multilevel mediation, MSEM does not require all outcomes to be measured at level-1, and can test effects of level-2 variables by modelling their effects on the level-2 part of level-1 variables; all in one analysis. The traditional ML-estimator is inappropriate due to unequal classroom sizes. The Muthèn's maximum-likelyhood-based estimator (MUML) is better suited to handle unbalanced data (Muthén, 1994).

The hypothesized model in this study was an 2-(1,1)-(1,1,1)-1 Upper-level mediation model, where the predictor Teacher Efficacy, was a level-2 measure, and all other variables were measured at level-1 . The model was tested following a procedure outline by Preacher, Zhang and Zyphur (2011). Their method is designed to build a MSEM by first determining the need to go multilevel, by calculating Intra-Class-Correlations (ICC's) of all variables. Generally, in large samples ($N > 100$), ICC's as low as .01 have been found to strongly inflate type I error rates (Barcikowski, 1981). The second step in the MSEM analysis is testing the fit of a model based on the within-variance of the data, which identifies covariance on the individual level. The model build based on the within-covariance matrix is then used in Step 3, where a model based on the between-variance is added. The model on the between level shows how level-2 variables interact with the other aggregated variables at level-2. In effect, two models are built and joined: a model at the individual student level, and a model at the classroom level.

To date there is no research available that has determined the appropriate sample size for mediation analysis in a MSEM framework. It has been suggested that 40 level-2 units are appropriate to detect large structural paths at the

between level, whereas N > 100 has been suggested to detect small effects (Meuleman & Billiet, 2009).

MSEM was performed on the first data set in EQS version 6.1. To interpret a model's fit, the following indicators were used: RMSEA and SRMR below 0.05 and CFI scores above 0.95 indicate good fit (Browne & Cudeck, 1992) and RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999).

4.3 Results

Prior to analysis, the research variables were examined for accuracy of data entry, missing values, and fit between their distributions. No variables had missing values over 5%, and there was no pattern to the missing data (MCAR's $\chi^2(79) = 94.98$, $p = .106$). Missing values were replaced by EM-estimates (Musil, Warner, Yobas, & Jones, 2002) based on all other research variables in the dataset. No continuous variables deviated from the normal distribution, and no univariate extreme cases (> 3*IQR) were found. 19 cases were identified through Mahalanobis distance as multivariate outliers ($p < .001$). Even though no specific pattern could be discerned, the cases were removed, leaving 1447 cases for analysis.

Table 4.1: Means, standard deviations, scale reliabilities and intraclass correlations of the background and research variables.

	N	**Mean**	**SD**	α	**ICC**
Teacher Efficacy	87	4.26	0.52	.77	–
Language proficiency	1447	4.43	0.65	.82	.06
Congruency monitoring	1447	-0.84	0.90	.91	.39
Congruency scaffolding	1447	-0.61	0.88	.87	.40
Interest	1447	2.98	0.97	.91	.27
Competence	1447	3.27	0.94	.86	.08
Relatedness	1447	3.38	0.83	.85	.28
Autonomy	1447	3.46	0.81	.67	.07

Table 4.1 shows means, standard deviations and the ICCs of the variables. The ICC of the variables ranged from .06 (Dutch language proficiency) to .40 (Congruency Scaffolding). These indicate that teacher variation can account for 6% to 40% of the variance of the variables. The results indicate that the data are not independent. MSEM is necessary for making valid statistical inferences.

The total covariance matrix was partitioned into pooled within- and between-sample covariance matrices. The square root of the ad hoc estimator constant or the scaling parameter was 4.30. The proposed model, analyzed on the within-sample covariance matrix, fit the data well, χ^2 (6) = 38.47, $p < .001$, N

= 1351, CFI = 0.99, SRMR = .03, RMSEA = 0.06. Not all parameters were significant, namely Congruency Monitoring with Competence, and Autonomy, and the covariance between, competence and relatedness, and autonomy and relatedness. Progressive removal of the non-significant links yielded a final within model that fit the data well, χ^2 (10) = 42.84, $p < .001$, N = 1351, CFI = 0.98, SRMR = .03, RMSEA = 0.05 (Figure 4.1).

Analysis of the between-variance matrix mirrored the final within model, but added the prediction of congruency in monitoring and scaffolding with teacher efficacy yielded a poor model fit, χ^2 (15) = 40.33, $p < .001$, N = 77, CFI = 0.90, SRMR = .14, RMSEA = 0.15. Adding the prediction of perceived competence and relatedness by teacher efficacy yielded an improved model ($\Delta\chi^2$ (2) = 20,34, $p < .001$), that fit the data reasonably well, χ^2 (13) = 19.99, p = .096, N = 77, CFI = 0.97, SRMR = .09, RMSEA = 0.08, considering the small sample size of between units.

The MSEM of the multilevel model on both the within and the between matrix fit the data well with χ^2 (30) = 54.25, p = .004, CFI = 0.99, SRMR = .048, RMSEA = 0.045. The multilevel model with parameter estimates and standard errors of parameter estimates are shown in Figure 4.1.

The model shows that while teacher efficacy for instructional strategies negatively corresponds with congruency in perceived AfL, the incongruency is somewhat negated by the positive association between teacher efficacy and student's perceived competence and relatedness. 33% of individual, and 70% of classroom level motivation could be explained by the model, which confirms the hypothesis that congruency in the perception of AfL-practice positively predicts students' intrinsic motivation, which is mediated by basic need fulfilment. High teacher efficacy and lower student language proficiency correspond with incongruent perceptions, which translate into lower intrinsic motivation.

The basic need for autonomy is a significant indicator for intrinsic motivation at the individual level, but not at the teacher level. However, the parameter weight is low in both levels. At the teacher level, teacher efficacy is a good indicator for classroom perceptions of competence and relatedness, offsetting the negative relation between teacher efficacy and congruency in AfL perceptions. At the individual level, language proficiency is a significant, albeit weak, indicator of congruent perceptions of AfL, whereas these relations are not significant at the teacher level.

Overall, at the individual level, the results show that student language proficiency corresponds to better congruency between teachers and students. Congruent perceptions of AfL, and congruency in perceived scaffolding in particular, positively relate to intrinsic motivation, mediated through basic need fulfilment. At the teacher level, the relation between congruent perceptions and classroom intrinsic motivation is less clear.

4.4 Discussion

Linking assessment perceptions to student motivation

The aim of this study was to test the hypothesis that differences between students and teachers in the perception of monitoring and scaffolding activities

Table 4.2: Tested within, between and MUML-multilevel models and their fit measures.

	Model	χ^2**(df)**	**CFI**	**SRMR**	**RMSEA**	$\Delta\chi^2$**(df)**
1	Within model_baseline	54.94*	.98	.03	.06	-
2	Within model parsimonious	55.20*	.98	.03	.05	0.26^{ns}(model 1 vs 2)
3	Between model	41.93*	.90	.13	.14	
4	Between model + efficacy -> comp, relat	20.53	.97	.08	.08	21,40*(model 5 vs 6)
5	MSEM (combined within, between)	68.68*	.98	.05	.05	

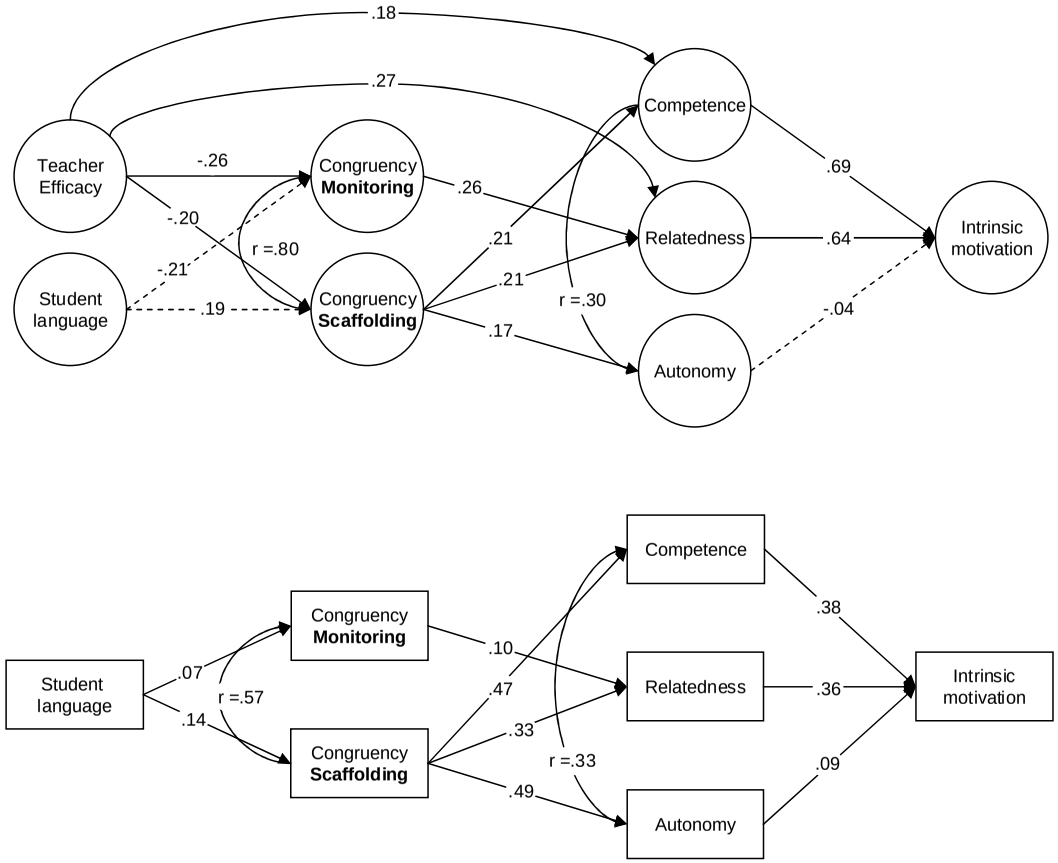


Figure 4.1: between (top) and within (bottom) model and parameter estimates. Nonsignificant lines are represented as dashed lines

would predict students' intrinsic motivation as mediated by the students' basic needs of Competence, Autonomy, and Relatedness. It was expected that students who are more proficient in the language of instruction would be more congruent with their teachers, and that teachers who are more efficacious in their teaching would be less congruent with their students.

Teacher-student congruency in AfL-perceptions and student motivation

The study's main hypothesis was confirmed; congruency in teacher-student perception of assessment for learning (AfL) practices, i.e., specifically monitoring and scaffolding, predicted higher student intrinsic motivation. In a multilevel structural equation model this effect accounted for 70% of classroom (aggregated group level) and 33% of individual students' intrinsic motivation.

Using the fit of the equation model, we were able to partly link congruency in perceptions of assessment practices on monitoring and scaffolding to students' language proficiency, and were able to establish a negative relation with teacher efficacy for instructional strategies. The negative teacher efficacy relationship with assessment for learning and student motivation could be explained by the argument that teachers who are highly confident in their ability to teach focus more on (stress and expect more of) student use of given feedback (monitoring or scaffolding) and highlight in their assessment for learning practice specifically their teaching goals and intentions, which subsequently may overestimate their students' evaluations and perceptions of AfL (Swann Jr. et al., 2007; Wheatly, 2002).

In clarifying our main expectation, that alignment in perceptions on assessment practices influences student motivation to learn, Self-determination Theory (SDT) (Deci & Ryan, 1985) was employed as a framework. The relationship between intrinsic motivation and congruency in the perception of Scaffolding was found to be mediated by basic need fulfilment, whereas the relationship between congruent teacher-student perceptions of Monitoring with interest was mediated by Relatedness only. The Framework of SDT also helps explain the (unexpected) effects found for teacher efficacy (Tschannen-Morann & Hoy, 1998). We found that high teacher efficacy negatively relates to congruency, but positively to fulfilment of the needs for competence and relatedness. Within the SDT framework we interpret this in the sense that efficacious teachers seem to inspire feelings of competence in their classrooms, and are perceived as more likeable, partly offsetting the negative effect of incongruent perceptions in their classrooms on classroom intrinsic motivation.

Some of our findings were at odds within the SDT framework. Autonomy, as a significant predictor of intrinsic motivation in our model at the individual level, seems to be a weak predictor at the classroom level. We found a large amount of within classroom variation, but very little at the classroom level for autonomy. While experiencing autonomy seems motivational for students, there seems to be little contribution from the classroom environment to those feelings of autonomy. It is possible that students do not strongly respond to teachers' efforts in supporting autonomy, or as Kunter and Baumert (2006) proposed after reporting similar findings: student ratings are easily influenced by their personal preferences. This is supported by our finding that there is a strong correlation between feelings of competence and feelings of autonomy,

which might indicate an interaction. Students who feel less competent might prefer more support, and thus less autonomy, whereas self-perceived competent students might appreciate the extra autonomy in their learning.

To summarize, the fit we found in our tested models for the relation between perceptions on assessment and student motivation: at the individual level, students who are proficient in the language of instruction are more aligned with their teachers in perceiving the assessment for learning practices in their classroom. Alignment in the perception of scaffolding activities such as clarifying goals and assessment criteria, and use of questioning coincides with high basic need fulfilment and subsequently, higher intrinsic motivation. Alignment in the perception of monitoring activities, such as providing feedback, and feedforward either through the teacher, or by supporting peer- and self-assessments coincides with high relatedness (friendly learning climate for the student). In classrooms with high efficacious teachers, however, there is less congruence in AfL perceptions between teacher and students which has a negative association with motivation. This was interpreted to be offset by the students' fulfilled needs for feeling competent, and relatedness to their more efficacious teacher.

The need for congruency in AfL perceptions

The results highlight the importance of a "fit" in perception on assessment meant to enhance learning: teachers' and students' mutual understanding on the nature and utilization of assessment information is a key in utilizing that information to enhance learning within an instructional context (Birenbaum et al., 2006). Perceived alignment in the intent and content of scaffolding and monitoring activities indicate clarity in goals and expectations between teachers and students on what needs to be learned and what progress in achievement is and needs to be made. The high amount of explained variance of congruency in assessment perceptions on motivation lends support to the claim that a mismatch results in a loss of effectiveness of instruction (Norman, 1986; Bartholomew et al., 2001). The issue of reaching congruency seems to be an important instructional activity, since we could relate it to instructional efficacy of teachers. We assume that it is likely that setting high expectations by teachers on goal attainment and setting standards for learning accomplishment of their students may be at odds with student perceptions of assessment practices meant to reflect their current mastery and to foster learning motivation. A high discrepancy in perceptions may be detrimental to the students' perception of being able to comply. The results indicate, unsurprisingly, that secondary school teachers in general have little influence on the language proficiency of their students, but that students low in language proficiency still have more difficulties in the understanding and recognition of AfL as support (Bartholomew et al., 2001). This would imply that interpersonal aspects of teachers' instructional efficacy may be more advantageous when teachers high in efficacy also take care to realize that language-ability differences complicate students' recognition of their support. A more statistical interpretation of the absence of classroom-language effects is that the already small effect of language at the individual level becomes even smaller when aggregated into classroom averages. The relatively small amount of classrooms to compare could have resulted in a loss of power.

Our findings lend support to the contention that assessment practices need to be carefully scrutinized as they are closely linked to motivation to learn. Especially with respect to the position teachers take in the delivery of assessment information by bridging the need of high expectations and student needs to be motivated to learn. Finding alignment according to the SDT framework we employed is established by addressing the competence and relatedness needs of the student.

Some drawbacks need to be mentioned however in suggesting too strong implications. Due to the cross-sectional nature of the data, it is not possible to draw causal conclusions about the nature of the relationship between teacher-efficacy and congruency of AfL perceptions, or about the possibility of changing congruency of perceptions by making teachers aware of the double edged sword of their own efficaciousness. Quasi-experimental research is needed to test the effectiveness of making teachers aware of the effects of their efficacy beliefs on their perceived practice and how it blocks their ability to adapt their teaching in order to keep their students motivated.

5

Ethnic differences in congruency of teacher-student assessment for learning perceptions and intrinsic motivation[4]

Abstract

Second generation immigrant students are often more intrinsically motivated for learning than native students. The aim of this study was to examine whether the relationship between perceptions of Assessment for Learning (AfL) and intrinsic student motivation differed between Dutch, Turkish and Moroccan students. It was expected that students who are more language-proficient would be more congruent with their teachers in perceiving AfL, and that teachers who are more efficacious in their teaching would be less congruent with their students. A total of 775 students in 58 classrooms from junior vocational high schools participated in this survey-study. Multi-group multilevel path modelling showed that the explanatory mediation model was invariant across ethnicities, and that students of second generation Turkish and Moroccan descent, despite their lower reported language proficiency, had closer congruency in AfL perceptions with their teacher than Dutch students, related more to their teachers, felt more competent, and were more intrinsically motivated.

5.1 Introduction

Many immigrant children and adolescents, supported by their parents, hold a firm idea that school is an important avenue to, or opportunity for social and economic mobility (P. Vedder & Horenzcyk, 2006). This is an important resource for immigrant youth as well as for their community. It is reflected in a school motivation that is generally at par or stronger than the school motivation of their national contemporaries (OECD, 2003). Student motivation is consistently found to positively impact student competency beliefs (Ames & Archer, 1988), perceived school-wellbeing (Kasser & Ryan, 2001), and is an important factor in preventing dropout (Legault et al., 2006). These are all positive consequences of a strong motivation which are particularly important for non-western immigrant students, because hitherto they insufficiently benefit from school: their academic performance is generally lower and rates of school drop-out are higher than of their national peers (OECD, 2003; Suárez-Orozco,

[4]This chapter is submitted for publication as: Pat-El, R.J., Vedder, P., Tillema, H.H., & Segers, M.S.R. (Under review). Ethnic differences in congruency of teacher-student assessment for learning perceptions and intrinsic motivation.

Suárez-Orozco, & Todorova, 2008). If students' motivation is actually such an important resource, what then might happen to it in schools, when nonwestern immigrant students are less capable of converting it into more successful school careers'.

Intrinsic motivation

The Self-Determination Theory of motivation proposes that students have to satiate "needs" to feel motivated for an educational activity (Deci & Ryan, 1985; Ryan & Deci, 2000). Intrinsic motivation is the tendency to engage in activities for the inherent joy an activity brings, and has been positively linked with persistence, mastery learning goals, deep learning, and well being (Ryan & Deci, 2000). Motivational needs consist, according to SDT, of a feeling of being autonomous, a sense of relatedness with others in the activity, and experiencing the competence to fulfill a given activity. These three needs are characterized by (Jang, 2008) as basic needs and by Chirkov (2009) as culturally invariant. If the basic premise holds that the SDT model of intrinsic motivation is culturally invariant (Chirkov, 2009), then it stands to reason that differing intrinsic motivation between ethnicities should be explained by differing need fulfillment and not also, as suggested by Den Brok (Den Brok, 2001), ethnic variation in their educational values, norms, or needs. Given the importance of the perceived learning environment in explaining intrinsic motivation through need fulfillment, it is expected that ethnic differences should be found either in the perception of the learning environment, or more likely, in the ethnic differences in the relative effects of the perceived learning environment on basic need fulfillment. Noels, Clèment and Pelletier (1999) stated that if students perceive the teachers' behavior as supportive and safe, they are more likely to accept their feedback (viz., Van Gennip, Segers, & Tillema, 2010). This led us to contend that provision of formative feedback might well affect intrinsic motivation of immigrant and national students differently.

Impact of perceptions of assessment on motivation

Formative feedback is considered a major tool in enhancing learning (Assessment Reform Group, 2002) but its impacts on students' motivation to learn is dependent on students' needs and past performance (Hattie, 2008; Ryan & Deci, 2000; Vollmeyer & Rheinberg, 2005). Feedback is formative when information is not only given to students as an indication of performance, but when it is used as an instrument to improve students' future learning, as well as the teacher's own teaching (P. Black & Wiliam, 1998a). Formative feedback has been characterized in the Assessment for Learning (AfL) literature by a small set of features which can be labeled as: monitoring and scaffolding (Pat-El, Tillema, et al., 2011). Monitoring provides the students with feedback information on current states of progress relative to the goals to be attained (i.e., 'where you are'- (Sadler, 2010)) and resembles what is known as 'giving knowledge of results' (Butler & Winne, 1995). Scaffolding is the process of supporting learners to pursue the next steps to enhance further learning by giving directions and advice (Shepard, 2005). Studies on the effects of either way of feedback provision on students' motivation show that both monitoring and scaffolding positively affect motivation (Corbalan, Kester, & Van Merriënboer,

2009; Shute, 2008). In particular, research on scaffolding shows that receiving extra information on how to improve on tasks had a positive influence on motivation (Dresel & Haugwitz, 2008; Shute, 2008). It has become clear that not only the provision of feedback or the perception by students of its provision is important in explaining student motivation, but also whether students and teachers mutually agree on whether and how AfL is practiced in classrooms (Pat-El, Segers, Tillema, & Vedder, 2011).

It is important that the student anticipates teacher instructions and feedback as personal expectations. Student expectations of the teacher's contribution prepare the student for more or less optimal 'absorption' or inclusion of the teacher provided information into the flow of learning. The less congruent teachers and students are in experiencing whether enough information is present to optimize instruction and learning, the less effective instruction and learning will be (Bartholomew et al., 2001; Doyle, 1977; Loughran, 2010; Skinner & Belmont, 1993).

Ethnic differences in perceptions of assessment and motivation

Many studies (e.g., Iyengar & Lepper, 1999; Kaplan & Maehr, 1999; Wong, Eccles, & Sameroff, 2003) found that immigrant students show marked differences in levels of intrinsic and extrinsic motivation compared to native peers, despite lower socio-economic status (Fuligni, 1997; Suárez-Orozco, Rhodes, & Milburn, 2009). It is of interest to identify why immigrants are more motivated in classrooms in the context of the support they receive on their learning (Massey, Gebhardt, & Garnefski, 2009; Wubbels, Den Brok, Veldman, & Van Tartwijk, 2006).

Earlier studies suggest that students do not perceive teacher behavior differently, but rather value it differently. The differential effect of AfL on student motivation is found to be related to the perceived teacher's classroom behavior (P. Black & Wiliam, 1998a; Entwistle & Tait, 1990; Wiliam, 2011). Differences in how perceptions of AfL impact student motivation might be culturally influenced due to cultural differences in (the interpretation of) teacher-student communication (Au & Kawakami, 1994). Congruency between teachers and students on AfL perceptions is more likely when students are proficient in the language of instruction (Pat-El et al., 2010). In the Netherlands, where the current study was conducted, immigrant students generally are less proficient in Dutch, the language of instruction, than their national contemporaries are (OECD, 2010). This could mean that immigrant students run a higher risk of misinterpreting teacher communication, and thus greater misalignment with their teachers. It is possible that despite lower language proficiency of immigrant students, differential effects of perceived teacher behavior might explain their higher intrinsic motivation

Student language proficiency is a likely candidate for explaining misalignments in perceived AfL. Another candidate is teacher efficacy, or teacher beliefs in being able to implement instructional strategies (Pat-El et al., 2010). High teacher efficacy was associated with a larger mismatch in AfL perceptions, which was by a possible self-verifying process (see Swann Jr. et al., 2007, for a review) in which teachers who are confident in their ability to teach might focus more on the feedback they give, and subsequently may overestimate their

students' evaluations of AfL. Even though high teacher efficacy might relate to greater misalignment (Gerges, 2001; Pat-El, Segers, et al., 2011; Pat-El et al., 2010; Wheatly, 2002), teacher efficacy is simultaneously positively linked to student motivation or motivational variables directly (Midgley, Feldlaufer, & Eccles, 1989; Pat-El, Segers, et al., 2011). Whether this complex relationship of teacher efficacy, perception misalignments, and motivation differs across ethnic populations is still unknown, however. High teacher efficacy has been found to positively relate to teacher persistence, resilience, attitude towards student errors (Ashton & Webb, 1986), and patience with students who struggle (Gibson & Dembo, 1984). Given the nature of this function as adaptive to students' needs it is likely that teacher efficacy, as a contextual factor, has a similar relationship for national, as well as immigrant students' perception of AfL practices.

The current study

This study examines whether differences in the effect of formative feedback on motivation for immigrants and native students can be interpreted from the mediating role of student motivational needs. The research questions that will be answered are: Are there ethnic differences in the perception and the effect of monitoring and scaffolding on students' intrinsic motivation mediated by the basic needs of competence, relatedness and autonomy? And, do either student proficiency in the language of instruction or teacher efficacy or both play a role in explaining possible differences? The theoretical model is presented in Figure 5.1.

5.2 Method

Sample

Students and teachers in junior vocational high schools in the Netherlands took part in this study. The individual students and their teachers were the unit of analysis. Questionnaires were administered in seven schools to 1466 students and 89 teachers. Classes were only included in the analysis whenever they included at least one second generation Moroccan or Turkish student. Students from ethnic groups that were too small to produce stable and reliable model estimates in MSEM were excluded from analyses. These included all first generation immigrants (16 in total), and students from, for example, Belgium, Surinam, and the Antillean islands. Eventually this resulted in a sample of seven schools with 775 students (N_{girls} = 386, N_{boys} = 355, missing = 34) and 58 teachers. The participating teachers represented a broad range of subjects: from arts to sciences. The average class size was 17.6 students (SD = 4.88; min = 11; max = 27). Median student age was 13 years (min = 11, max = 18). Over 70% (72.9%; N = 565) of the participants were of Dutch ethnicity, 10.8% (N = 84) were second-generation Moroccan immigrants, and 16.3% (N = 126) were second generation Turkish immigrants.

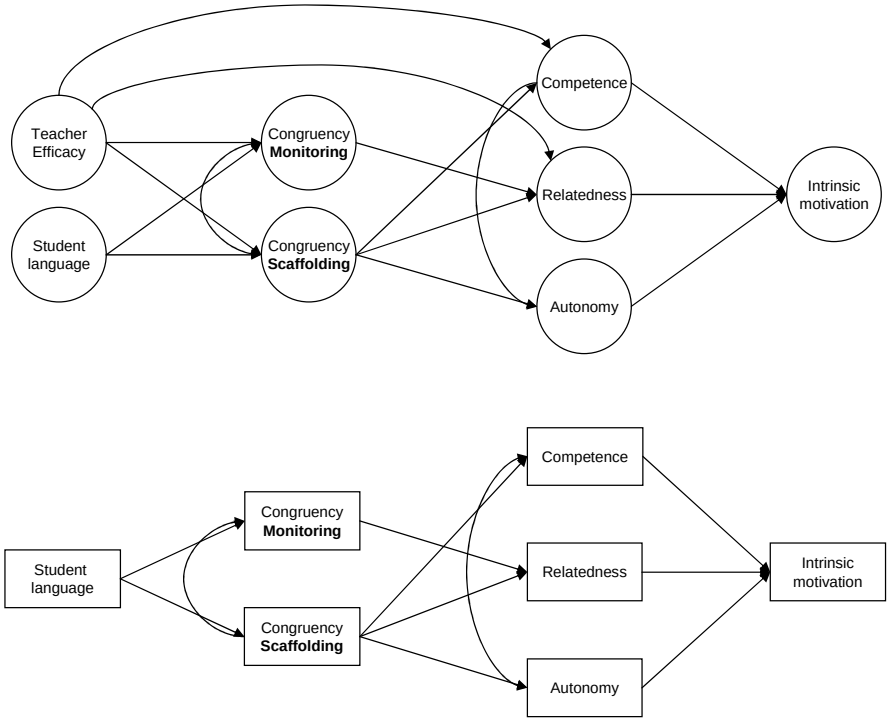


Figure 5.1: Theoretical multilevel model

Measures

Questionnaires were used to measure AfL perceptions in teachers and students in conjunction with teachers' efficacy for teaching, students' language proficiency, three basic needs for motivation, and their intrinsic motivation. All Likert items response values ranged from 1 (strongly disagree) to 5 (strongly agree), unless otherwise specified.

Intrinsic motivation.

Intrinsic motivation was measured with the interest/enjoyment scale from the Intrinsic Motivation Inventory (IMI) (McAuley, Duncan, & Tammen, 1989). The scale was translated to Dutch and adapted to measure interest and enjoyment in the class they were being taught at that moment. The scale consisted of 7 Likert scale items. Sample items are: "I would describe this class as very interesting" and "I think this is a boring class" (reversed). In the current study, Cronbach's α of the 7-item scale was .91.

Perceived competence.

The 4-item Perceived Competence Scale (Deci et al., 1981) was translated to Dutch. A sample item is: "I am capable of learning the material in this class." In the current study, Cronbach's α of the 4-item scale was .86.

Relatedness.

Relatedness was measured with the 8-item Relatedness scale from the IMI (Deci et al., 1981), which was translated to Dutch and adapted to measure the class they were being taught at that moment. A sample item is: "I feel like I can really trust my teacher". In the current study, Cronbach's α of the 4-item scale was .85.

Perceived autonomy.

The 3-item Perceived Autonomy Scale (Martens & Kirschner, 2004) was in Dutch. A sample item is: "I can determine for myself how I work during this class." In the current study, Cronbach's α of the 3-item scale was .67.

Independent variables

Perception of AfL congruency.

Perceptions of AfL practices were measured with the Assessment for Learning Questionnaires which consist of both a teacher (TAFL-Q) and a student version (SAFL-Q) (Pat-El, Tillema, et al., 2011). The questionnaire is comprised of two subscales: Monitoring and Scaffolding. All items were scored on 5-point Likert scale items. Differences (i.e., alignment) were computed by subtracting the teacher scores, from the TAFL-Q, from the student scores, from the SAFL-Q. Negative difference scores therefore indicate higher teacher scores, while positive difference scores indicate higher student scores. Items in both the TAFL-Q and the SAFL-Q are worded similarly to enable strict comparisons between the two populations. The Monitoring subscale consisted of 12 items (Cronbach's $\alpha = .91$). The scale was defined in terms of frequency and form of feedback and how self-monitoring is facilitated. Sample items are "I encourage my students to reflect upon how they can improve their assignments," and "I discuss with my students how to utilize their strengths to improve on their assignment."

The Scaffolding subscale consisted of 16 items (Cronbach's $\alpha = .87$). The scale refers to communication of clear learning goals and criteria, and how those goals and criteria can be met. Sample items are "I ensure that my students know what they can learn from their assignments," and "I adjust my instruction whenever I notice that my students do not understand a topic."

Ethnicity.

Ethnicity was a categorical variable and defined by country of birth of the student and their parents: when both student and parents were born in The Netherlands, the student's ethnicity was coded native Dutch. Immigrant students in the sample had very diverse ethnic backgrounds. To allow for meaningful comparisons among groups, only the largest ethnic populations in the

sample, with sufficient N, were selected for analysis (viz. second generation Turkish and Moroccan students). Students were coded as second generation immigrants when they were born in The Netherlands, but at least one parent was born in another country.

Teacher efficacy for instructional strategies.

Teachers rated their efficacy for teaching on the Ohio State Teacher Efficacy Scale (OSTES) (Tschannen-Morann & Hoy, 1998). The 4-item Efficacy for instructional strategies was translated into Dutch. A sample item is "To what extent can you provide an alternative explanation or example when students are confused" Cronbach'sα in the present study was .77.

Dutch language proficiency.

Student language proficiency was measured with a self-report scale from the ICSEY study (Berry et al., 2006). Self-reports to determine language proficiency have a high correlation with evaluations of a person's language proficiency by others (Kirchmeyer, 1993). In the four-item scale students were asked to evaluate how well they were able to read, write, speak, and understand the Dutch language. Scores ranged from 1 (not at all) to 5 (very well). Cronbach's α in the present study was .82.

Procedure

Schools were randomly selected across the Netherlands by email and telephone. Of the 31 schools that were approached, 7 agreed to participate in this study. Teachers and their students participated by informed consent. The questionnaire was filled out during one arbitrarily selected course hour in the presence of research assistants, which took teachers and students about 25 minutes. Respondents were assured that their contribution was anonymous. Students received a small reward when they returned their fully completed questionnaire.

Analysis

Multigroup Multilevel Structural Equation Models (MSEM) (Muthén, 1994) were used to compare the fit of the hypothesised model to the different ethnic groups in Dutch schools. The Muthèns maximum-likelyhood-based estimator (MUML) was used, because of its better handling of unbalanced data in contrast to the traditional ML-estimator (Muthén, 1994). The hypothesized model in this study was an 2-(1,1)-(1,1,1)-1 Upper-level mediation model, where the predictor Teacher Efficacy, was a level-2 measure, and all other variables were measured at level-1 . The model was tested following a procedure outlined by Preacher, Zhang and Zyphur (2011). Their method is designed to build a MSEM by first determining the need to go multilevel, by calculating Intra-Class-Correlations (ICC's) of all variables. Generally, in large samples ($N >$ 100) ICC's as low as .01 have been found to strongly inflate type I error rates (Barcikowski, 1981). The second step in the MSEM analysis is testing the fit of a model based on the within-variance of the data, which identifies correlations on the individual level. The model build based on the within data is then used in Step 3, where a model based on the between-variance is added. The

model on the between level shows how level-2 variables interact with the other aggregated variables at level-2. In effect, two models are built and joined: a model at the individual student level, and a model at the classroom level.

To date there is no research available that has determined the appropriate sample size for mediation analysis in a MSEM framework. It has been suggested that 40 level-2 units are appropriate to detect large structural paths at the between level, whereas N > 100 has been suggested to detect small effects (Meuleman & Billiet, 2009).

MSEM was performed on the first data set in MPlus version 7. To interpret a model's fit, the following indicators were used: RMSEA and SRMR below 0.05 and CFI scores above 0.95 indicate good fit (Browne & Cudeck, 1992) and RMSEA and SRMR below .08 and CFI scores above .90 indicate acceptable fit (Hu & Bentler, 1999). CFI is penalized for model complexity, which means that in complex models CFI might drop. A measure that does not penalize large or complex models is the Gamma (γ) (Fan & Sivo, 2007), which is calculated based on the number of manifest variables, df, and RMSEA, and should have values above .90 for acceptable fit and above .95 for good fit.

Multigroup testing was applied to the tested models to assess structural invariance of the questionnaires between ethnic groups, by evaluating the invariance of variable intercepts between groups (Gregorich, 2006). A critical value of -0.01 ΔCFI against a configural invariant model will be used to judge invariance (Cheung & Rensvold, 2002). The constraints which are not helpful for the model are released. The final fit of the basic, similar model and that of the separate models for the two groups, are compared on differences.

5.3 Results

Prior to analysis the research variables were examined for accuracy of data entry, missing values, and fit between their distributions. No variables had missing values over 5%, and there was no pattern to the missing data (MCAR's $\chi^2(7) = 8.525$, $p = .289$). Missing values were replaced by EM-estimates (Musil et al., 2002) based on all other research variables in the dataset. No continuous variables deviated from the normal distribution, and no univariate extreme cases (> 3*IQR) were found. Eleven cases were identified through Mahalanobis distance as multivariate outliers ($p < .001$). As no specific pattern could be discerned, besides combinations of very high and very low scores on several variables, the cases were retained, leaving 775 cases for analysis. Table 1 shows means, standard deviations, and the ICCs of the variables for each ethnic group. ICC ranged from .06 (Morrocan intrinsic motivation) to .48 (Dutch congruency in Monitoring). ICC overall are fairly similar across groups, with the exception of Intrinsic motivation which is highest for the Dutch students (ICC = .33), and lower for the Turkish (ICC = .19) and Moroccan students (ICC = .06). The high ICCs indicate that data are not independent, and that MSEM is necessary for making valid statistical inferences.

Table 5.1: Means, standard deviations and intra class correlations for the total sample and the subsamples.

	Total		**Dutch**			**Moroccan**			**Turkish**		
	M	**SD**	**M**	**SD**	**ICC**	**M**	**SD**	**ICC**	**M**	**SD**	**ICC**
Intrinsic motivation	3.05	0.97	2.94	0.97	0.33	3.24	1.02	0.06	3.43	0.83	0.19
congruency	-0.84	0.95	-0.93	0.94	0.48	-0.69	1.01	0.48	-0.54	0.86	0.46
Monitoring congruency	-0.60	0.85	-0.64	0.81	0.36	-0.68	1.04	0.44	-0.39	0.88	0.46
Scaffolding											
Competence	3.33	0.96	3.24	0.92	0.07	3.58	1.11	0.12	3.56	0.92	0.11
Relatedness	3.51	0.77	3.48	0.77	0.34	3.43	0.80	0.30	3.68	0.77	0.45
Autonomy	3.46	0.89	3.49	0.86	0.10	3.33	1.02	0.16	3.42	0.93	-0.07
Language proficiency	4.39	0.66	4.46	0.62		4.23	0.91		4.18	0.59	
Teacher efficacy	4.25	0.53	4.21	0.55	-	4.31	0.48	-	4.25	0.56	-

Note: Dutch $N = 565$, Moroccan $N = 84$, Turkish $N = 126$

Ethnic invariance of Congruency of AfL perceptions on intrinsic motivation.

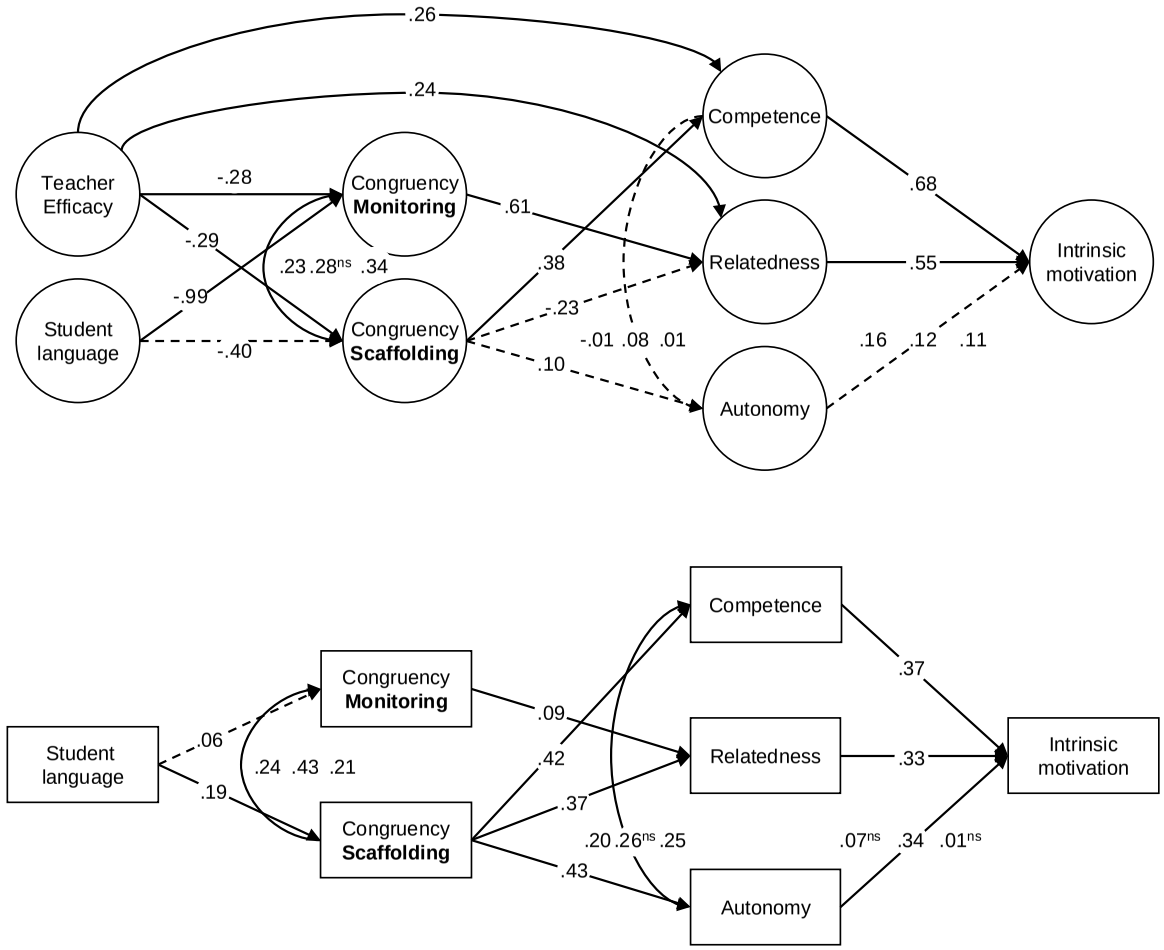
The hypothesized model was tested in Mplus version 7, with all paths and intercorrelations constrained to be equal across the populations. The multilevel-multigroup model was found to be homogeneous across the three populations, $\chi^2(118) = 175.59$; CFI = .95, γ = .94; SRMR_W = .03, SRMR_B = .09, RMSEA = .05, except for the prediction of intrinsic motivation through autonomy, and the intercorrelations between monitoring and scaffolding, and between competence and autonomy. Autonomy's prediction of intrinsic motivation was only significant and strong for the second generation Moroccans. All parameter estimates were invariants across Dutch Turkish and Moroccan students, except for the path between perceived autonomy and intrinsic motivation. The effect between autonomy and intrinsic motivation was significant in the Dutch and Turkish sample, but not in the Moroccan. The model and parameter estimates are summarized in Figure 5.2.

Ethnic differences in AfL congruency, basic needs and intrinsic motivation.

Comparisons between ethnic groups with Tukey HSD corrections showed ethnic differences on key variables, as summarized in Table 1. In agreement with the reported literature and as expected, second generation Turkish (M = 3.05, SD = 0.97, r = .07) and Moroccan students (M = 3.24, SD = 1.02, r = .15) were more intrinsically motivated than Dutch students (M = 2.94, SD = 0.92). Contrary to our expectations, second generation Turkish (M = -0.54, SD = 0.86, r = .21) students held more teacher-congruent perceptions of Monitoring than Dutch students (M = -0.93, SD = 0.94), and more teacher-congruent perceptions of Scaffolding (M = -0.39, SD = 0.88) than Moroccan (M = -0.68, SD = 1.04, r = .15) and Dutch students (M = -0.64, SD = 0.81, r = .15), despite Dutch students (M = 4.46, SD = 0.62) higher reported ability in the Dutch language than second generation Turks (M = 4.18, SD = 0.59, r = .23) and Moroccans (M = 4.23, SD = 0.91, r = .15). In terms of basic needs Turks (M = 3.68, SD = 0.77, r = .13) felt more related to their teacher than Dutch (M = 3.48, SD = 0.77) students. Turkish (M = 3.56, SD = 0.92, r = .17) together with Moroccan students (M = 3.58, SD = 1.11, r = .16) felt more competent than Dutch students (M = 3.24, SD = 0.92). There were no significant differences between the three ethnic groups on perceived autonomy.

5.4 Discussion

The aim of this study was to examine ethnic differences in the relationship between perceptions of Assessment for Learning and student intrinsic motivation. It was expected that ethnic differences in student-teacher agreement on AfL-practice mediated by fulfillment of the three basic needs of motivation, feelings of competence, relatedness and autonomy, would help explain possible differences between native Dutch students and the largest minority groups in the Netherlands.



model.pdf

Figure 5.2: Multilevel multigroup model and parameter estimates. Not-invariant parameters have been split in order of Dutch, Moroccan, Turkish estimates respectively. Dotted lines were not significant in all groups.

The results confirm most of the study's expectations. Second generation Turkish and Moroccan students were more intrinsically motivated than native Dutch students. The theoretical model of how teacher-student congruency in the perception of AfL relates to intrinsic motivation, mediated by basic need fulfillment, was invariant across groups. At the teacher level, the relationship between teacher efficacy, perception congruencies, and basic need fulfillment was equal across the three sampled ethnic groups. The absence of differential effects, except for the relationship between autonomy and intrinsic motivation, between the studied populations implies that this model of motivation is not interactional and that second generation Turkish and Moroccan pupils are very similar to Dutch students in how the variables are interrelated. Differences in motivation could be explained on the basis of differences in the independent variables: Turkish students have a stronger agreement in perceptions relative to their teacher, then Dutch and Moroccan pupils, higher relatedness with their teacher, and a higher intrinsic motivation. Even though the Turkish students report a stronger agreement in AfL perceptions, the Moroccan students' intrinsic motivation is still higher, which might be explained by the stronger relationship between autonomy and intrinsic motivation. Still, our findings are not completely in line with the expectations of this study. It was expected that students who are less proficient in the Dutch language agree less with their teacher about how much AfL is practiced in the classroom, and that this would apply for the second generation immigrants. While we do find that the second generation immigrants report a lower proficiency in the Dutch language, their perceptions of AfL are more congruent with those of their teachers. Given that within the group of second generation Turkish students language still has a negative relation to AfL perception-congruency, which was invariant in relation to the Dutch students, it is likely that some unmeasured confounding factor has a competing, and stronger impact on AfL congruency. It is suggested by Den Brok and colleagues (Den Brok, Tartwijk, Wubbels, & Veldman, 2010) that teacher-student interpersonal relationships are more important for students with a non-Dutch background than for students with a Dutch background, and more important for second generation than for first-generation immigrant students. In our study this differential effect was not replicated, as the relationship between relatedness to the teacher and intrinsic motivation was invariant in our study. These differences might be explained by the different focus: in the study by Den Brok and colleagues' (2010) study, the independent factor was interpersonal teacher behavior (classroom management and harmony of interactions). It is likely that the teachers' interpersonal behavior is differently perceived by students of other ethnicities (Den Brok et al., 2003), which might moderate the relationship between the students' perception of the learning environment and experienced relatedness to the teacher.

Limitations

This study could not fully confirm Chirkov's (2009) finding that the Self-determination theory of basic needs predicting intrinsic motivation was invariant across ethnicities. The relationship between perceived autonomy and intrinsic motivation was not invariant for Dutch students, second generation Turkish and Moroccan students. It would be hasty, however, to dismiss Chirkov's

(2009) findings; even though the Cronbach's α for the perceived autonomy scale of .67 is acceptable, it still is rather low for a validated scale. Given the small sample size of Moroccans (N= 84) for path analysis purposes it remains unclear whether lack of invariance was due to power problems or because of theoretical reasons. More focused research is needed to verify the universality of the Self-Determination theory.

Although significant, the effect sizes of the differences between ethnicities range from very weak to weak. Despite the theoretical relevance, the differences are small, and it begs the question: in what way ethnic differences in intrinsic motivation can be addressed by targeting the problem of student-teacher perception incongruencies. It remains unclear whether more pronounced differences are to be expected in different populations of immigrants. It is to be expected that first generation immigrants, and/or immigrants who have not yet acculturated into the host society should differ more from their national peers (Den Brok et al., 2010), but this question would need to be researched with representative samples of both first and second generation immigrants.

Implications

Notwithstanding the study's limitations, and whether ethnic mean differences are considered large enough to be relevant, the results still underline the importance of awareness of cultural differences in responsiveness to classroom practices. If AfL is to integrate assessment with learning and foster student motivation, it is important that students' and teachers' perceptions on the nature and content of the assessment provided are congruent (Bartholomew et al., 2001; Loughran, 2010; Norman, 1986; Pat-El, Segers, et al., 2011). The positive message is that although levels of congruency and intrinsic motivation may vary between ethnic groups, the interrelations do not. Teachers need not be culturally sensitive, but they need to be sensitive and adaptive to student differences irrespective of students' cultural background.

Future research should both try to provide clarifications for the cultural differences in motivation and the role AfL perceptions can play, and in what way educators can be trained to take account of these differences.

6

General Discussion

Feedback is one of the strongest predictors of students' motivation for learning (Hattie & Timperley, 2007), but needs to be accepted in order to be successful (P. Black & Wiliam, 2009). Assessment for Learning (AfL) is a two-way process in which not only students adapt their learning with information provided by assessments, but teachers, when they experience a lack of understanding or learning in students, adapt their teaching as well. In this respect, it is of great importance that students' and teachers' perceptions on the nature and content of the assessment provided are congruent. Both need to know where to focus on to further learning and instruction. The aim of this thesis was to explore whether teachers and students are in agreement about the level to which AfL is practiced in the classroom, and whether perception congruencies help explain intrinsic motivation in the classroom.

AfL congruency between teachers and students.

To help answer the research question of this dissertation, about the congruency of assessment for learning perceptions between teachers and students, we developed two instruments to gauge AfL: the SAFL-Q for students and the TAFL-Q for teachers (Chapter 2). Two dimensions of AfL were found to describe the generally agreed upon principles of Assessment for Learning. The first, scaffolding, refers to classroom interaction wherein learning goals and criteria are clarified through and in addition to classroom questioning. The second, monitoring, refers to analyzing student learning progress with the intention to find challenges and opportunities for optimizing learning. The matching items between the two versions and the strong level of measurement invariance mean that, in quantitative analyses of congruency, the different scores between teachers and students can be similarly interpreted for both teachers and students alike.

As expected, student-teacher incongruencies were large. Incongruencies for both monitoring and scaffolding varied greatly between classrooms, were highly homogeneous within classrooms (Chapter 3), and students' perceptions were consistently more negative than those of their teachers. These outcomes are similar to the results of earlier studies, such as those from MacLellan (2001) and Schultz (2001) who showed that teachers also tended to report more use of formative assessment and instructional strategies than their students. As reported in a study by Könings (2007) teachers often indicate that students are invited to take an active part in the instruction and course design, whereas students report they are not.

Do teacher-student congruent perceptions predict intrinsic motivation?

The study's main hypothesis was confirmed; congruency in teacher-student perception of assessment for learning practices (i.e., monitoring and scaffolding) predicted higher student intrinsic motivation (Chapter 4). In a multilevel structural equation model this effect accounted for 70% of classroom (aggregated group level) and 33% of individual students' intrinsic motivation. In accordance with Self-Determination Theory (Ryan & Deci, 2000), the congruency of monitoring and scaffolding perceptions predicts students' intrinsic motivation. The relationship between intrinsic motivation and congruency in the perception of scaffolding was found to be fully mediated by the fulfilment of students' basic needs for competence, autonomy and relatedness, whereas the relationship between congruent teacher-student perceptions of monitoring with interest was mediated by relatedness only. The reason to focus on intrinsic motivation in particular, rather than on intrinsic motivation in addition to extrinsic motivation, lies in earlier research findings. The Self-Determination Theory has been found to accurately describe the process of becoming intrinsically motivated (e.g., Krapp, 2005), but similar validations of how differential fulfilment of self-determination theory's basic needs of motivation affect the more extrinsic forms of motivation (Ryan & Deci, 2000) are still lacking. The relationship between the two is not well understood and its study is hampered by conceptual as well as instrumental challenges (Birenbaum et al., 2006; Hattie & Timperley, 2007).

The question remains in what way perceptual congruency between teachers and students is a more important predictor of intrinsic motivation than solely the perspective of the student. The results of this dissertation offer an insight into the complex relation between perceptions and classroom factors. The reported cross-sectional association between perception congruencies and motivational factors, such as relatedness towards the teacher (Chapter 3), support our hypothesis and confirm conclusions or hypotheses of earlier studies (Loughran, 2010; MacLellan, 2001). Now that it has become apparent that congruent perceptions associate positively with need fulfillment and intrinsic motivation, it would be of particular interest to confirm this hypothesis in experiments to determine whether congruencies are important causes of intrinsic motivational gains.

However, designing an experiment to test this hypothesis is daunting, given the highly interpersonal nature of the interactions involved in the learning environment. This interpersonal nature is suggested throughout this dissertation, where large intra-class correlations indicate a strong within-class 'agreement' between students about the level of congruence in perception between teachers and students as to the use of AfL. Whereas the ICC provides helpful insight in the tendency for (in)congruencies to be present across classrooms as a whole, instead of being purely individual differences, the data does not let us infer in what way student perceptions are dependent of other students in the classroom. It may very well be that perceptions or evaluations of teachers converge towards agreement over time, not only because of shared experiences of teaching practice, but also because of the emergence of some kind of common consensus or opinion between students about their teachers' qualities.

The way intrinsic motivation of students is associated with AfL-perception

congruency was found to be similar for Dutch students and the largest minority groups in the Netherlands, namely second generation Moroccan and Turkish immigrant students (Chapter 5). The results confirm most of the study's expectations. Second generation Turkish and Moroccan students were more intrinsically motivated and the Turkish students in particular showed stronger agreement in perceptions relative to their teacher and higher relatedness than native Dutch students. The absence of differential effects, except for the relationship between autonomy and intrinsic motivation, between the studied populations implies that this model of motivation is not interactional and that second generation Turkish and Moroccan pupils are very similar to Dutch students in how the variables are interrelated. Differences in motivation could be explained on the basis of differences in the independent variables.

Student Language proficiency and congruency of AfL perceptions.

Closer alignment of teacher and student perceptions was found to be related to better student language proficiency (Chapters 3 through 5). The importance of language in understanding the learning environment was underscored by Vygotsky (1978) who highlighted that through practical activity a child constructs meaning on an intrapersonal level, while speech connects this meaning with the interpersonal world shared by the child and her/his culture. The association of language proficiency with scaffolding congruency was stronger than with monitoring congruency. This is especially relevant given the importance of scaffolding for basic needs fulfilment and intrinsic motivation. It might well be that scaffolding activities, which emphasize the need for communication of clear goals and criteria, demands more of teachers and students in terms of being able to recognize subtleties in teacher-student communication than monitoring.

The effect of language proficiency seems robust across datasets, but the strength of the association is rather subdued. Based on the data presented in this dissertation it is difficult to conclude that the association is weak. The self-report measure of Language proficiency showed strong ceiling effects, which may have resulted in loss of power, because of the resulting low variance. There are a number of non self-report language proficiency tests that are either used by educational assessors, or educational researchers (Aarts & Kurvers, 2001)(see Bachman, 2000, for a review), which might be good additions to the self-report scale used in this study for cross reference. The role of language proficiency is probably better tested through triangulation of instruments, rather than one test, to rule out unfamiliarity with testing situations, or a one sided emphasis on specialized skills and components of language

In addition, it might be worthwhile to focus more on particular dimensions of language proficiency. There is a subtle, but clear distinction between academic or school and learning bound language and common, every-day language. In academic language, words may be more difficult than in common language and often carry a more specific definition. Sentences tend to be longer and more complex, and the subjects discussed are oftentimes more difficult and more abstract than usually discussed in common language, for instance at home (Chaudron, 1988; Schleppegrell, 2001). Formative feedback is an example of a type of interaction that students most likely only will experience at

school in which predominantly academic language is used; and is regarded as a different language proficiency (Cummins, 1981). In as much as this complicates the analysis of the role of language, it also helps in developing a more general focus on students, independent of their ethnic background or generational status. The focus on ethnic background is converted into a focus on language performance.

The contradictory nature of Teacher-efficacy on AfL perceptions and motivation.

Teacher efficacy has consistently been found to positively (either directly or indirectly) associate with key student outcomes, such as student motivation (Midgley et al., 1989), student self-efficacy (Anderson, Greene, & Loewen, 1988) student achievement and openness to new ideas (Allinder, 1994). Other studies, however, suggest that teacher efficacy might also block variation in employing instructional strategies (Gerges, 2001), module use, and course organization (Wheatly, 2002). A similar pattern of contradictory findings was found in the studies presented in this dissertation (Chapters 3 through 5). In line with the theory of self-verification (Swann Jr. et al., 2007), which proposes that people would rather learn something about themselves which confirms their self-perceived desirable qualities, than their undesirable qualities, efficacious beliefs about instructional design, and to a lesser degree about student engagement, predicted less congruence between teachers' and students' perception of AfL. We found strong effects for teacher efficacy as regards instructional strategies and, to a lesser extent, teacher efficacy with respect to student engagement. Both engaging students and utilizing varied assessment strategies are core elements of AfL (Assessment Reform Group, 2002) and it would seem likely that if any self-verification processes were taking place, they would relate to efficacy for instructional strategies and student engagement rather than for efficacy related to classroom order. That efficacy for maintaining classroom order did not associate with perception congruency in this dissertation, adds evidence to our theory that a self-verification process is likely to take place. Using Self-determination theory, the contradictory nature of teacher-efficacy could be modeled: high teacher efficacy negatively relates to congruency of both monitoring and scaffolding, but positively relates to fulfilment of the needs for competence and relatedness, which mediate the relationship with intrinsic motivation. Efficacious teachers seem to inspire feelings of competence in their classrooms, and are perceived as more likeable, partly offsetting the negative effect of incongruent perceptions in their classrooms on classroom intrinsic motivation.

Limitation and challenges

Several limitations of the studies reported in this dissertation give rise to questions that need to be addressed in future research.

Multilevel nature of the research

In Chapters 3 to 5, multilevel analyses were conducted to reduce as much as possible the impact of the hierarchically nested nature of the data. In Chapter

2 we did not take the multilevel nature into account, although we did use a nested data structure. Due to the decision to maintain the chapter as close as possible to the version as it was published in the British Journal of Educational Psychology (Pat-El, Tillema, et al., 2011), the study was not amended to a multilevel validation study. Even though a multilevel study would add to the quality of the validation, the current validation study is more than satisfactory in providing evidence for a good fit of the proposed factorial solution, as usually multilevel analysis increases the fit of confirmatory factor models instead of decreasing it (Byrne, 2006). This is not a guarantee though, and future studies are needed to confirm the dimensionality of the AfL construct, and the validity of the questionnaire not only at the individual level, but also at the class level.

Difference scores

This dissertation provided support for the existence and importance of differences between teacher and student perceptions in the classroom. However, the use of difference scores remains a contested issue (Cronbach, 1958; Edwards & Parry, 1993). Difference scores are said to be less reliable than more direct forms of comparisons, but this claim is not widely supported, and even regarded as a myth (Edwards, 2001). Nevertheless, difference scores are challenged and this is in part because of the difficulty in interpreting relative distances between groups. This is especially relevant in this dissertation, because individual student perceptions are compared to general teacher perceptions. Those perceptions might be inflated, because teachers can perceive themselves as avid providers of feedback whenever they practice it, even for a few students. As long as other students have other experiences, a class-level mismatch emerges, without any real way of determining whether teachers would agree with those individual students when asked about them specifically. Another problem can be the reliability of scales at their endpoints. Even though tests of homoscedasticity did not reveal any pattern of varying levels of variance, scales do tend to become less sensitive at their extreme points. Even though latent intercepts were found to be invariant between teachers and students in Chapter 2, it is still clear that teachers score at the high end of the AfL-scales. It is possible that the lower reliabilities of the teacher scales, as evidenced by their lower Cronbach's alpha, might stem from teachers extreme scores. This could have resulted in an undetermined impact on any difference scores derived of them.

There are few solutions to the problem of difference scores. Relevant alternatives have methodological issues of their own, such as using teacher and student scores separately and model an interaction effect. This approach does not represent effects of congruence and at best yields approximations that become similar to difference scores (Edwards, 2001). Other alternatives, such as polynomial regression offer more reliability, but do nothing to alleviate the basic problem of comparing general scores to individual scores.

The results of this study generate at least one other difference-score based hypothesis: the difference itself seems predictive of student motivation, regardless of whether this difference occurs on the high end or relatively low end of the perceived level of practise for both teachers and students. It is possible that the difference itself is responsible for the association with student intrinsic motivation. Unfortunately the present study was not designed to adequately

test this hypothesis. Teachers in general scored high on both AfL subscales, and actual practice was not empirically observed. This dissertation does not provide evidence that mismatches have similar effects for teachers who report low levels of AfL practice.

Future studies might focus on more direct measures of comparison so that analyses truly compare how teachers perceive their own feedback behaviour towards individual students. To evaluate whether congruency in perceptions is itself positively related to motivation, irrespective whether the congruency is about low or high AfL-practice, it would be important to select teachers a-priori who are more traditional in their teaching style. By selecting low-AfL practising teachers, the full congruency spectrum can be scrutinized.

Cross-sectional design

The present study provided support for the hypotheses that perceptual differences between teachers and students as regards the quality of teacher support and feedback are associated to student motivation. However, the cross-sectional nature of the analysis and data-gathering preclude the possibility of causal attributions. To further test the hypotheses formulated and tested in this dissertation, (quasi-) experiments are needed. As stated earlier, designing experiments to test the hypotheses of this dissertation is daunting. However, they are a necessity for justifying attempts to translate the finding of the role of perceptual differences in student motivation into classroom interventions, or teacher education.

The importance of congruent AfL perceptions on intrinsic motivation

The studies reported in this thesis yielded four important outcomes. First, students' and teachers' perceptions of the level to which AfL is practised in classrooms are largely incongruent. Teachers perceive more practice of AfL than students. Second, these incongruencies are strongly associated with intrinsic motivation, explained by students' lower feelings of personal competence and autonomy to learn in their own way, and less interpersonal feelings of relatedness towards their teacher. Third, teachers' positive impact on motivational interpersonal factors is partly offset by their own possible over-efficaciousness, which might limit the degree to which they reflect on their own teaching. This relationship seems to hold for students regardless of their ethnic background. Moreover, variation in intrinsic motivation between Dutch, Moroccan and Turkish students seem to be explained by their differing levels of congruence with their teachers' perceptions of AfL practice. Fourth, teacher-efficacy associates with intrinsic motivation in a dualistic fashion: high teacher efficacy negatively relates to congruence, but positively to fulfilment of the needs for competence and relatedness. Even though teachers may be conducive to the maintenance and emergence of feelings of competence in their classrooms, and are perceived as having good relationship with their students, there is also a negative relation between efficacy and the perception of congruencies. These findings are in line with arguments in the literature. Teacher efficacy is positively related to student outcomes, such as motivation (Ashton & Webb, 1986), but teacher efficacy may also interfere with teachers ability to vary in

their instruction (Gerges, 2001; Wheatly, 2002). The implications of these results highlight the need to instruct teachers to reflect on their own teaching. AfL is for the large part dependent on the teacher to be able to adapt their own teaching (Wiliam, 2011), and future research should focus on determining whether helping teachers to reflect on their work helps in creating closer teacher-student perception congruencies, while maintaining the positive effects of efficacy on motivational needs.

7

Nederlandse samenvatting

De motivatie van leerlingen voor het leren op school hangt samen met hoe er wordt lesgegeven. Om gemotiveerd te raken zijn er voor leerlingen, volgens de zelfdeterminatietheorie van Ryan en Deci, drie psychologische behoeften die bevredigd dienen te worden: een gevoel dat de leerling in staat is te slagen (competent voelen), een mate van zelfbeschikking ervaren, en daarmee vrijheid om het leren op een eigen wijze uit te voeren (autonoom voelen), en een persoonlijke 'klik' met de docent ervaren (verbondenheid). In dit proefschrift is onderzocht hoe de ervaring van de leeromgeving zich verhoudt tot deze drie basisbehoeften om zo de motivatie te verklaren.

Motiverende leeromgevingen zijn leeromgevingen waar de nadruk ligt op het leerproces in tegenstelling tot de beoordeling. Onderwijs bezien vanuit een sociaal-constructivistische invalshoek stoelt op de aanname dat leerlingen het beste leren als zij zelf en met elkaar actief betrokken zijn bij processen van kennisconstructie. De docent heeft in een sociaal-constructivistische leeromgeving meer de rol van ondersteuner van het leerproces, dan van kennisoverdrager. Regelmatige (informele) evaluaties kunnen als eikpunten dienen voor zowel leerlingen, die informatie krijgen over te verbeteren punten, als voor leerkrachten, die inzicht krijgen waar aan gewerkt moet worden in volgende lessen. Wanneer dergelijke evaluaties informeel en regelmatig worden toegepast, dan sluit dit aan bij een leerbenadering die Evalueren om te Leren (Assessment for Learning) wordt genoemd. Bij Evalueren om te Leren wordt benadrukt dat leren vooral motiverend is als voor leerlingen duidelijk is wat de leerdoelen en beoordelingscriteria zijn, waardoor ze zelfstandig kunnen overzien en begrijpen wat ze nog moeten doen om de leerdoelen te bereiken.

Recent blijkt dat niet zozeer de leeromgeving op zich een belangrijke voorspeller is voor leermotivatie, maar hoe de leeromgeving wordt waargenomen. Dit geldt niet alleen voor leerlingen. Ook hoe leerkrachten hun lesgeven ervaren blijkt van invloed te zijn op hun lesgeven. Zoals reeds besproken is dit lesgeven weer van invloed is op de motivatie van leerlingen. Hieruit voortvloeiend kunnen percepties van leerlingen en leerkrachten elkaar aanvullen bij het bepalen van de leermotivatie van leerlingen. Echter de invloeden van beide percepties hoeven niet optimaal positief of cumulatief te zijn. Het blijkt dat leerkrachten en leerlingen de lespraktijk verschillend ervaren. Leerkrachten menen hun leerlingen niet teveel te belasten, hen veel eigen ruimte te bieden en leerdoelen helder uit te leggen, terwijl leerlingen vaak juist een hoge werkdruk, weinig beslissingsruimte en veel onduidelijkheid over de leerdoelen ervaren. Deze incongruenties kunnen voor wrijving zorgen tussen leerkracht en leerling en mo-

gelijk tot problemen leiden rond de leermotivatie van individuele leerlingen, of zelfs van meerdere of zelfs alle leerlingen in een klas.

Eerder onderzoek verduidelijkt dat tweede-generatie immigranten jongeren meer gemotiveerd zijn dan hun autochtone Nederlandse leeftijdsgenoten. Uit eerder onderzoek is tevens gebleken dat verscheidenheid qua culturele achtergrond van leerlingen tot uiting komt in verschillende beoordeling van of waardering voor het gedrag van leerkrachten. Het is dan de vraag of de hogere motivatie van deze groep leerlingen te verklaren valt door de mate waarin de percepties van het lesgeven van leerlingen en leerkrachten bij elkaar aansluiten. Hierover is weinig bekend. Dit is een belangrijke vraag in deze dissertatie.

In dit proefschrift is onderzocht in welke mate congruentie tussen leerlingen en leerkrachten in de beleving van Evalueren om te Leren verband houdt met de motivatie van leerlingen. In de conceptualisering en operationalisatie van motivatie laten we ons met name leiden door de zelfdeterminatietheorie.

In Hoofdstuk 2 is allereerst een meetinstrument ontwikkeld en gevalideerd om de verschillende ervaringen van Evalueren om te Leren tussen docenten en leerlingen te kunnen vergelijken. Evalueren om te Leren werd gedefinieerd in twee dimensies: monitoring en scaffolding. Monitoren behelst de docentfeedback (wat ging goed of niet goed en hoe wordt het beter), maar ook de inzet van de docent om leerlingen te leren zichzelf te kunnen evalueren. Scaffolding verwijst naar instructies en discussie in de klas en de mate waarin leerdoelen en beoordelingscriteria expliciet worden gemaakt. De leerling- (SAFL-Q) en leerkrachtversie (TAFL-Q) van de vragenlijst gebruiken zo identiek mogelijke bewoordingen. In een confirmatieve factor-analyse is de vragenlijst voor leerlingen eerst gevalideerd. Deze bleek structureel meetinvariant te zijn met de versie voor de leerkrachten. Dit maakt het mogelijk om op kwantitatieve wijze de ervaring van monitoring en scaffolding te meten en zonder vertekening de leerlingen met de leerkrachten te vergelijken in hun beleving van Evalueren om te Leren.

In Hoofdstuk 3 is onderzocht of leerlingen en leerkrachten verschillen in hun beleving van Evalueren om te Leren in de klas en welke factoren als verklaringsgrond kunnen dienen voor de individuele verschillen. Er werd verwacht, op basis van de theorie van zelfverificatie, dat leerkrachten die veel vertrouwen hebben in hun eigen bekwaamheid voor lesgeven (teacher-efficacy) voornamelijk de positieve signalen uit de klas opvangen en die als bevestiging zien van het positieve beeld dat ze van zichzelf hebben. Op het niveau van de leerlingen werd verwacht dat door het belang van communicatie, en daarmee het talige karakter van Evalueren om te Leren, het niveau van de Nederlandse taalvaardigheid van leerlingen ook verband zou houden met de mate van incongruentie tussen leerling en docent in de beleving van Evalueren om te Leren. Deze verwachtingen werden bevestigd. Voor zowel monitoring als scaffolding waren er grote verschillen tussen de belevingen van leerkrachten en hun leerlingen: docenten rapporteerden structureel meer inzet van Evalueren om te Leren dan hun leerlingen dat deden. Met name docenten met een hoge mate van vertrouwen in hun bekwaamheid om verscheidene instructiestrategien toe te passen bleken meer incongruent met hun leerlingen. Dit gold ook voor leerlingen met een mindere, zelfgerapporteerde, Nederlandse taalbeheersing.

In Hoofdstuk 4 wordt een onderzoek gerapporteerd waarin is onderzocht in welke mate congruente beleving van Evalueren om te Leren tussen docenten en leerlingen samenhangt met voor leerlingen hogere motivatie voor het leren

en of dit verband met motivatie kon worden verklaard aan de hand van de zelfdeterminatietheorie. Dit bleek het geval te zijn. Een hogere congruentie in de beleving van scaffolding ging samen met een sterkere motivatie. Dit verband kon worden verklaard door de mate waarin leerlingen meenden dat hun behoeften aan autonomie, verbondenheid en competentie werden bevredigd. De samenhang tussen congruentie in de beleving van monitoring en motivatie, kon verklaard worden door gevoelens van verbondenheid met de docent. Een ander opvallende bevinding betrof het oordeel van leerkrachten over hun bekwaamheid om instructiestrategien toe te passen. Conform bevindingen in Hoofdstuk 2 hing een sterker gevoel van bekwaamheid samen met meer incongruente belevingen. Tegelijkertijd correspondeerden sterkere gevoelens van bekwaamheid bij de leerkrachten met sterkere gevoelens van competentie en verbondenheid bij leerlingen. Het lijkt erop dat hoewel docenten in hun beoordeling van onderwijsleersituaties relevante informatie missen, waardoor er een grotere incongruentie met leerlingen ontstaat, zij tegelijkertijd aardiger worden gevonden door hun leerlingen, die tevens meer van oordeel zijn dat ze competent zijn om te leren.

In Hoofdstuk 5 werd getoetst in welke mate het onderzochte model van zelfgedetermineerde motivatie vanuit congruentie van de beleefde leeromgeving kon helpen verklaren dat Marokkaans-Nederlandse en Turks-Nederlandse leerlingen sterker gemotiveerd zijn dan de Nederlandse leerlingen. Het voorgestelde model, bevestigd in Hoofdstuk 4, kon dezelfde verklaringsgrond bieden voor Nederlandse als Marokkaanse en Turkse jongeren. De modellen waren in belangrijke mate invariant, behalve de wijze waarop ervaren autonomie motivatie voorspelt. Dit verband is sterker voor de Marokkaanse leerlingen dan voor de Nederlandse en Turkse leerlingen. Zoals verwacht waren de Turkse en Marokkaanse leerlingen meer gemotiveerd dan de Nederlandse leerlingen. Zij ervoeren een hogere mate van verbondenheid met hun leerkracht en hun belevingen van Evalueren om te Leren waren meer congruent met hun docenten dan die van de Nederlandse leerlingen.

In deze dissertatie is gewezen op het belang van congruentie in percepties van leerkrachten en leerlingen over Evalueren om te Leren. Die congruentie blijkt afhankelijk van de Nederlandse taalvaardigheid van leerlingen en van het beeld dat leerkrachten van zichzelf als professional hebben. Etnische verschillen blijken weliswaar een rol te spelen in de niveaus van leermotivatie, in de waardering voor bepaalde aspecten van leersituaties en in de Nederlandse taalvaardigheid, maar de modellen ter verklaring van congruentie tussen leerlingen en leerkrachten en verschillen in leermotivatie gelden voor alle leerlingen. Voor leerkrachten lijkt het belangrijk dat ze zich ervan bewust worden dat er incongruente belevingen van leersituaties kunnen bestaan met leerlingen, maar ook op welk gebied en waardoor ze ontstaan. Ze kunnen incongruentie voorkomen of verminderen door deze kennis in te zetten bij observaties van en gesprekken met leerlingen. Hierdoor zouden leerlingen meer zelfvertrouwen en autonomie kunnen ervaren, en een sterkere persoonlijke band met de leerkracht kunnen beleven, waardoor hun motivatie om zich op school in te zetten, zou kunnen toenemen.

8

References

References

Aarts, R., & Kurvers, J. (2001). Ouders, taal en interactie in Opstap Opnieuw [Parents, language and interaction in the 'Opstap Opnieuw' program]. *Toegepaste Taalwetenschap in Artikelen*, *66*(2), 9–22.

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, *17*, 86–95.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*, 260–267.

Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers and students thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research*, *34*, 148–165.

Argyris, C. (1977). Organizational learning and management information systems. *Accounting, Organizations and Society*, *2*(2), 113–123.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York, NY: Longman.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles.* Available from `http://arg.educ.cam.ac.uk/CIE3.pdf`

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, *72*(2), 261–278.

Au, K. H., & Kawakami, A. J. (1994). Cultural congruence in instruction. In E. R. Hollins, J. E. King, & W. C. Hayman (Eds.), *Teaching diverse populations: Formulating a knowledge base* (pp. 5–23). Albany, NY: State University of New York Press.

Bachman, L. (2000). Modern language testing at the end of the century: Assuring that what we need counts. *Language testing*, *17*(1), 1–42.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: Freeman.

Barcikowski, R. S. (1981). Statistical power with the group mean as the unit of analysis. *Journal of Educational and Behavioral Statistics*, *6*, 267–285.

Bartholomew, L. K., Parcel, G. S., Kok, G. S., & Gottlieb, N. H. (2001). *Intervention mapping: Designing theory and evidence-based health promotion programs.* CA: Mayfield: Mountain View.

Berry, J. W., Phinney, J. S., Sam, D. L., & Vedder, P. H. (2006). Immigrant youth: Acculturation, identity, and adaptation. *Applied Psychology: An International Review*, *55*, 303–332.

Biggs, J. B. (1990). Effects of language medium of instruction on approaches to learning. *Educational Research Journal*, *5*, 18–28.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. S. R. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., et al. (2006). A learning integrated assessment system. *Educational Research Review*, *1*, 61–67.

Black, A. E., & Deci, E. L. (2000). The effects of instructors autonomy support and students autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*, 740–756.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, *21*, 49–97.

Black, P., & Wiliam, D. (1998b). Inside the black box - Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–149.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5–31.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale*, *54*(2), 199–231.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley & Sons.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*, 151–167.

Boyle, J. P. (1987). Sex differences in listening vocabulary. *Language Learning*, *2*, 273–284.

Broadfoot, P. (1998). Records of achievement and the learning society: A tale of two discourses. *Assessment in Education*, *5*, 447–477.

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students.* New York, NY: Nova Science Publishers.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230–258.

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

Byrne, B. M. (2006). *Structural equation modeling with EQS.* Mahwah, NJ: Erlbaum.

Carver, C. S., & Scheier, M. F. (2000). On the structure of behavioral self-regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 41–84). San Diego, CA: Academic.

Central Bureau of Statistics. (2007). *Jaarboek onderwijs in cijfers [Yearbook of education in numbers].* Heerlen, The Netherlands: Central Bureau of Statistics.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* Cambridge, UK: Cambridge University Press.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Chirkov, V. I. (2009). A cross-cultural analysis of autonomy in education : A self-determination theory perspective. *Theory and Research in Education*, *7*, 253–262.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, *33*, 509–531.

Corbalan, G., Kester, L., & Van Merriënboer, J. J. G. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, *19*(6), 455–465.

Cornelius-White, J. (2007). Learner-centred teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, *77*, 113–143.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200.

Cronbach, L. J. (1958). Person perception and interpersonal behavior. In R. Tagiuri & L. Petrullo (Eds.), (pp. 353–379). Stanford, CA: Stanford University Press.

Crooks, T. (2001). *The validity of formative assessments.* Leeds, UK: British Educational Research Association.

Cummins, J. (1981). *The role of primary language development in promoting educational success for language minority students.* Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning.* Council of Chief State School Officers (CCSSO). Available from `http://www.ccsso.org/publications/details.cfm?PublicationID=381`

Daugherty, R., Black, P., Ecclestone, K., James, M., & Newton, P. (2007). *Investigating the alignment of assessment to curriculum.* London, UK.

Davis, E. A. (2006). Characterizing productive reflection among preservice elementary teachers: Seeing what matters. *Teaching and Teacher Education*, *22*, 281–301.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self determination in human behavior.* New York, NY: Plenum Press.

Deci, E. L., Schwartz, A. J., Sheinman, L., & Ryan, R. M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology*, *73*, 642–650.

Den Brok, P. (2001). *Teaching and student outcomes: A study on teachers thoughts and actions from an interpersonal and a learning activities perspective.* Unpublished doctoral dissertation, Utrecht University, Utrecht, the Netherlands.

Den Brok, P., Levy, J., Wubbels, T., & Rodriguez, M. (2003). Cultural influences on students perceptions of videotaped lessons. *International Journal of Intercultural Relations*, *27*, 355–374.

Den Brok, P., Tartwijk, J. van, Wubbels, T., & Veldman, I. (2010). The differential effect of the teacherstudent interpersonal relationship on student outcomes for students with different ethnic backgrounds. *British Journal of Educational Psychology*, *80*(2), 199–221.

Dochy, F. J. R. C., & McDowell, L. (1997). Introduction: Assessment as a Tool for Learning. *Studies in Educational Evaluation*, *23*(4), 279–298.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education*, *5*, 163–198.

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *The Journal of Experimental Education*, *7*, 3–18.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*(6), 1577–1613.

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment: Introduction to the special issue. *Higher Education*, *22*, 201–204.

Entwistle, N. J., & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education*, *19*, 169–194.

Eren, A. (2009). Self-efficacy beliefs and mental time travel ability: Uncovering a hidden relationship in educational settings. *Learning and Individual Differences*, *19*, 513–517.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*, 509–529.

Fuligni, A. J. (1997). The academic achievement of adolescents from immigrant families: The roles of family background, attitudes, and behavior. *Child Development*, *68*, 351–363.

Gerges, G. (2001). Factors influencing preservice teachers variation in use of instructional methods: Why is teacher efficacy not a significant contributor? *Teacher Education Quarterly*, *28*, 71–88.

Gibbs, G., & Simpson, C. (2003). Measuring the response of students to assessment: The Assessment Experience Questionnaire. In *11th international improving student learning symposium.* Hinckley, UK.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569–582.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment.* London, UK: Falmer Press.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? . *Medical Care*, *44*, S78–S94.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, *4*, 63–69.

Hall, B., Burley, W., Villeme, M., & Brockmeijer, L. (1992). *An attempt to explicate teacher efficacy beliefs among first year teachers.* San Fransisco.

Harlen, W., & Crick, R. D. (2003). Testing and Motivation for Learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169–207.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 Meta-analyses relating to achievement.* London, UK: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students conceptions of assessment: Factorial and structural invariance of the SCoA across sex, age, and ethnicity. *European Journal of Psychological Assessment*, *25*, 30–38.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the role of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, *76*, 349–366.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Curriculum Journal*, *17*, 109–138.

Jang, H. (2008). Supporting students motivation, engagement, and learning during an uninteresting activity. *Journal of Educational Psychology*, *100*, 798–811.

Jang, H., Reeve, J., Ryan, R. M., & Kim, A. (2009). Can Self-Determination Theory Explain What Underlies the Productive, Satisfying Learning Experiences of Collectivistically Oriented Korean Students? *Journal of Educational Psychology*, *101*(3), 644–661.

Järvelä, S., Lehtinen, E., & Salonen, P. (2000). Socioemotional orientation as a mediating variable in teaching learning interaction: implications for instructional design. *Scandinavian Journal of Educational Research*, *44*, 293–306.

Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, *74*, 443–471.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits-self-esteem, generalized self-efficacy, locus of control, and emotional stability with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*, 80–92.

Kaplan, A., & Maehr, M. L. (1999). Achievement goals and student well-being. *Contemporary Educational Psychology*, *24*, 330–358.

Kasser, T., & Ryan, R. M. (2001). Be careful for what you wish: Optimal functioning and the relative attainment of intrinsic and extrinsic goals. In P. Schmuck & K. M. Sheldon (Eds.), *Life goals and well-being: Towards a positive psychology of human striving* (pp. 116–131). Goettingen, Germany: Hogrefe & Huber.

Keiny, S. (1994). Constructivism and teachers' professional development. *Teaching and Teacher Education*, *10*(2), 157–167.

Kirchmeyer, C. (1993). Multicultural task groups: An account of the low contribution level of minorities. *Small Group Research*, *24*, 147–148.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Könings, K. D. (2007). *Student perspectives on education: Implications for instructional redesign.* Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen, the Netherlands.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction*, *15*, 381–395.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling.* London, UK: Sage Publications.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251.

Legault, L., Green-Demers, I., & Pelletier, L. (2006). Why do high school students lack motivation in the classroom? Toward an understanding of academic amotivation and the role of social support. *Journal of Educational Psychology*, *98*, 567–582.

Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Educational Psychology*, *96*, 68–84.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, *31*(3), 313–327.

Loughran, J. J. (2010). *What expert teachers do: Enhancing professional knowledge for classroom practice.* Sydney, London: Allen & Unwin/Routledge.

MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education*, *26*, 307–318.

Martens, R., & Kirschner, P. (2004). *How many factors predict intrinsic motivation? A preliminary exploration.* Utrecht, the Netherlands.

Massey, E. K., Gebhardt, W. A., & Garnefski, N. (2009). Self-generated goals and goal process appraisals: Relationships with sociodemographic factors and well-being. *Journal of Adolescence*, *32*(3), 501–518.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58.

McMillan, J. H. (2007). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction* (4th ed.). Boston, MA: Pearson/Allyn & Bacon.

Meuleman, B., & Billiet, J. (2009). A monte carlo sample size study: How many countries are needed for accurate multilevel SEM. *Survey Research Methods*, *3*(1), 45–58.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247–258.

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, *24*(7), 815–829.

Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research*, *22*(3), 376–398.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85–95.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-

regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*, 199–218.

Noels, K., Clément, R., & Pelletier, L. (1999). Perception of teachers communicative style and students intrinsic and extrinsic motivation. *The Modern Language Journal*, *83*, 24–33.

Norman, D. A. (1986). *Cognitive engineering.* Hillsdale, NJ: Lawrence Erlbaum Associates.

OECD. (2003). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003* (Tech. Rep.). Paris, France.

OECD. (2005). *Formative assessment: Improving learning in secondary classrooms* (Tech. Rep.). Paris, France.

OECD. (2010). *Policy Review of Migrant Education for the Netherlands* (Tech. Rep.). Paris, France. Available from `www.oecd.org/edu/migration/countrynotes`

Olkinuora, E., & Salonen, P. (1992). *Adaptation, motivational orientation and cognition in a subnormallyperforming child: A systematic perspective for training* (B. Wong, Ed.). New York, NY: Springer-Verlag.

Oortwijn, M. B., Boekaerts, M., & Vedder, P. (2008). The impact of teacher's role and pupils' ethnicity and prior knowledge on pupils' performance and motivation to cooperate. *Instructional Science*, *36*(5), 251–268.

Pat-El, R. J., Segers, M. S. R., Tillema, H. H., & Vedder, P. H. (2008). Constructing a new assessment for learning questionnaire. In *fourth biennial EARLI/Northumbria Assessment Conference.* Pottsdam, Germany.

Pat-El, R. J., Segers, M. S. R., Tillema, H. H., & Vedder, P. H. (2011, August). Teacherstudent perceptions of assessment practices as predictors of student motivation to learn. In *14th biannual conference of the European Association for Learning and Instruction.* Exeter, UK.

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2011). Validation of Assessment for Learning questionnaires for teachers and students. *British Journal of Educational Psychology*.

Pat-El, R. J., Tillema, H. H., Segers, M. S. R., & Vedder, P. H. (2010, September). Explaining mismatching perceptions of assessment for learning practices in lower vocational education. In *5th biennial EARLI/Northumbria Assessment Conference.* Hexham, UK.

Popham, J. (2008). *Transformative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevl SEM. *Structural Equation Modeling*, *18*, 161–182.

Prosser, M., & Trigwell, K. (1993). Development of an approaches to teaching questionnaire. *Research and Development in Higher Education*, *15*, 468–473.

Raviv, A., Raviv, A., & Reisel, E. (1990). Teachers and students: Two different perspectives?! Measuring social climate in the classroom. *American Educational Research Journal*, *27*, 141–157.

Reeve, J., Bolt, E., & Cai, Y. (1999). Autonomy-supportive teachers: How they teach and motivate students. *Journal of Educational Psychology*, *91*, 537–548.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosen-

berg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*, 151–161.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.

Sadler, R. D. (2010). Indeterminacy in the use of preset criteria for assessment in grading in higher education. *Assessment and Evaluation in Higher Education*, *34*, 159–179.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, *23*(4), 391–402.

Schleppegrell, M. J. (2001). Linguistic features of the language of school. *Linguistics and education*, *12*(4), 431–459.

Schultz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instructions and corrective feedback. *Modern Language Journal*, *85*, 244–258.

Scriven, M. (2002). Evaluation ideologies. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation: Viewpoints on educational and human services evaluation models* (2nd ed., pp. 249–278). Boston, MA: Kluwer Academic Publishers.

Segers, M. S. R. (2004). *Assessment en leren als een twee-eenheid: Onderzoek naar de impact van assessment op leren.* Leiden, the Netherlands: Leiden University.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, *63*(3), 66–70.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to succes in standards-based schools. *Phi Delta Kappan*, *87*(4), 324–328.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, *30*, 325–341.

Suárez-Orozco, C., Rhodes, J., & Milburn, M. (2009). Unraveling the immigrant paradox academic engagement and disengagement among recently arrived immigrant youth. *Youth Society*, *41*, 151–185.

Suárez-Orozco, C., Suárez-Orozco, M. M., & Todorova, I. (2008). *Learning a new land: Immigrant students in American society.* Cambridge: MA/London, UK: Harvard University Press.

Swann Jr., W. B., Chang-Schneider, C., & Angulo, S. (2007). Self-verification in relationships as an adaptive process. In J. Wood & A. Tesser (Eds.), *Self and relationships* (pp. 49–72). New York, NY: Psychology Press.

Tschannen-Morann, M., & Hoy, A. W. (1998). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783–805.

Van der Slik, F. W. P., Driessen, G., & De Bot, K. L. J. (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudal

multilevel examination in Dutch. *European Sociological Review*, *22*, 208–293.

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290.

Van Grinsven, L., & Tillema, H. H. (2006). Learning opportunities to support student self-regulation: Comparing different instructional formats. *Educational Research*, *48*, 77–91.

Vedder, P., & Horenzcyk, G. (2006). Acculturation and the school context. In D. L. Sam & J. W. Berry (Eds.), *Psychology of acculturation: International perspectives* (pp. 419–438). Cambridge, UK: Cambridge University Press.

Vedder, P. H. (1985). *Cooperative learning: A study on processes and effect of cooperation between primary school children.* Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen, the Netherlands.

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, *15*(589-602).

Vygotsky, L. (1978). *Interaction between learning and development.* Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1997). Narrative tools of history and identity. *Culture & Psychology*, *3*, 5–20.

Wheatly, K. F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education*, *18*, 5–22.

Whitelock, D. (2010). Activating Assessment for Learning: are we on the way with Web 2.0? . In M. J. W. Lee & C. McLoughlin (Eds.), *Web 2.0-based-e-learning: Applying social informatics for tertiary teaching* (pp. 319–342). IGI Global.

Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Fransisco, CA: Jossey-Bass.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*, 3–14.

Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, *99*, 181–193.

Wong, C. A., Eccles, J. S., & Sameroff, A. (2003). The influence of ethnic discrimination and ethnic identification on African American adolescents school and socioemotional adjustment. *Journal of Personality*, *71*, 1197–1232.

Worthington, A. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assessment and Evaluation in Higher Education*, *27*, 49–64.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment Research and Evaluation*, *12*, 1–25.

Wubbels, T., Brekelmans, M., & Hooymayers, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, *8*(1), 47–58.

Wubbels, T., Den Brok, P., Veldman, I., & Van Tartwijk, J. (2006). Teacher

interpersonal competence for Dutch secondary multicultural classrooms. *Teachers and Teaching*, *12*, 407–433.

Biography

Ron Pat-El was born on July 26, 1980 in Leidschendam, the Netherlands. He started his academic career at the 'Onze Blijde Wereld'-primary school in Voorburg at age 5. At age 12, he started his secondary school career at the Dalton-college in Voorburg and ended it successfully at College 't Loo in Voorburg. In 2000, he started studying Psychology at Leiden University from which he received his master's-degree in 2006 in Social and Organizational Psychology with an additional specialization in Statistics and Research Methodology. His thesis was about achievement-motivation of women in relation to perceived stigmatization of women at work. During his studies, he tutored in statistics and performed with his accordion. In the final stages of his studies, he found student- and research-assistant positions to fulfil. His PhD research at the Institute of Educational Studies at Leiden University focused on the motivation of pre-vocational high-school students and how it is influenced by the correspondence between student and teacher perceptions of assessment practices. His research interests include perceptions of feedback and assessment practices, social-cognitive classroom processes and classroom diversity.