This MS is the final prepublication (open access) version of the published article:

Marsh, Kuyper, Seaton, Parker, Morin, Möller, & Abduljabbar. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*(4), 326-341.
This article may not exactly replicate the final published version in the journal. It is not the copy of record and readers are encouraged to obtain the copy of record through their university or local library using the article's DOI (digital object identifier).

Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation

Abstract

In a comprehensive study (15,356 Dutch 9th grade students from 651 classes in 95 schools) we empirically tested the dimensional comparison theory (DCT) propositions formulated by Möller & Marsh (2013) as an extension of I/E theory, exploring methodological, theoretical, and substantive insights. According to DCT, academic self-concepts (ASC) are formed in relation to dimensional comparisons in different school subjects, as well as to social and temporal comparisons. In support of DCT predictions, paths from achievement to ASC in matching domains were substantially positive, but paths to non-matching domains (e.g., math achievement to verbal self-concept) were significantly negative. Extending DCT, we show that the more dissimilar the subjects, the more negative the cross paths (far comparisons), whereas cross paths relating more similar subjects (near comparisons) are much less negative and sometimes positive. Extending previous self-concept research and its integration with DCT, we found that positive paths for matching domains and negative paths for non-matching domains were larger for class marks based on classroom performance than for standardized test scores. Controlling for direct measures of social comparison (meVclass ratings of how each student compares to classmates) substantially reduced positive paths from achievement to ASC in matching domains, but also reduced the size of the negative paths from non-matching domains. Supplemental analyses suggest that dimensional comparison processes in both subjective rankings and actual class marks are consistent with those found in ASCs.

Keywords

Academic self-concept Dimensional comparison theory Social comparison theory Frame of reference effects CFA/SEM

1. Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation

Self-concept is one of the oldest constructs in psychology, a major focus in many disciplines, and an important mediating factor that facilitates the attainment of various desirable outcomes aside from positive self-concept itself (Marsh, 2007). In educational settings, a positive academic self-concept (ASC) is both a highly desirable goal and a means of facilitating subsequent academic achievement, academic accomplishments, and educational choice behaviors such as subject choice, coursework selection, academic persistence, and long-term educational attainment (e.g., Chen et al, 2013, Guay et al, 2004, Marsh, 1991, Pinxten et al, 2010, Parker et al, 2013). Theoretical models of ASC formation underscore the importance of frames of reference (Marsh, 2007): The same objective achievements can lead to highly different self-concepts, depending on the standards of comparison or frames of reference that individuals use to evaluate themselves, and can have important implications for future choices, performance, and behaviors.

In the broader psychological literature, the two most frequently posited frames of reference are social and temporal comparisons (Albert, 1977, Festinger, 1954, Möller, 2005, Möller et al, 2009, Möller et al, 2011); self-perceptions are based in part on how current accomplishments compare with past performances (temporal comparisons) and how they compare with the accomplishments of others in one's immediate context (social comparisons; e.g., classmates in one's school or class). However, in their theoretical founding of dimensional comparison theory (DCT), Möller and Marsh (2013) emphasize that: "Although social comparison (Festinger, 1954) and temporal comparison (Albert, 1977) theories are well established, dimensional comparison is a largely neglected yet influential process in self-evaluation" (p. 544). DCT (Marsh et al, 2014, Möller, Marsh, 2013) incorporates the extensive body of educational psychology research based on the I/E model, placing dimensional comparisons into a broader theoretical foundation in relation to more general psychological models of self-evaluation, person perception, frames of reference, and social comparison. In one of the first empirical studies based on the newly expanded DCT, the objectives of the present investigation are to provide:

1. empirical research specifically designed to test new theoretical predictions based on DCT and its extension of the classic I/E model;

2. the integration into DCT of existing self-concept research and new theoretical predictions about the distinct predictive effects of class marks (i.e., school grades on report cards) and standardized test scores on ASC; and

3. new applications of meVclass ratings (how my achievement compares with those of others in my class, globally and in specific school subjects), proposed by Huguet et al. (2009) as pure measures of social comparison into DCT, thereby more clearly separating the social and dimensional comparison predictive effects that are central to DCT.

2. Dimensional comparison theory (DCT): extension of the internal/external frame of reference (I/E) model

2.1. Theoretical basis of the original I/E model

The I/E model (Marsh, 1986) was originally developed to provide a theoretical basis to explain why math and verbal ASCs (MSC and VSC) are almost uncorrelated, even though academic achievements in the corresponding areas are substantially correlated (for further discussion, see Marsh, 2007). The theoretical processes posited in the I/E model are that ASC in a particular school subject is formed in relation to an external (social comparison) reference in which students compare their perceptions of their own performances in a particular school subject with the performances of other students in the same school subject, and an internal (dimensional, ipsative comparison) reference in which students compare their own performance in one school subject with their own performances in other school subjects. Thus, students may have a favorable MSC if math is their best subject, even though they are not particularly good at math relative to other students. The joint operation of these theoretical processes, depending on the relative weight given to each, is consistent with the near-zero correlation between MSC and VSC, which led to the revision of the Shavelson, Hubner, and Stanton (1976) multidimensional, hierarchical model of self-concept (see Marsh, 2007).

In empirical tests of theoretical predictions based on the I/E model (Marsh, 1986), MSC and VSC are regressed on math and verbal achievements (see Fig. 1A). Theoretically, the external comparison process predicts that good math skills lead to higher MSCs and that good verbal skills lead to higher VSCs. According to the internal dimensional comparison process, however, good math skills lead to lower VSCs once the positive effects of good verbal skills are controlled: The better I am at mathematics, the poorer I am at verbal subjects, relative to my good math skills. Similarly, better verbal skills lead to lower MSCs once the positive effects of good math skills are controlled. In models used to test these theoretical predictions (see Fig. 1A), the horizontal paths leading from math achievement to MSC and from verbal achievement to VSC (matching paths) are predicted to be substantially positive, but the cross paths leading from math achievement to VSC and from verbal achievement to MSC (Fig. 1) are predicted to be negative.

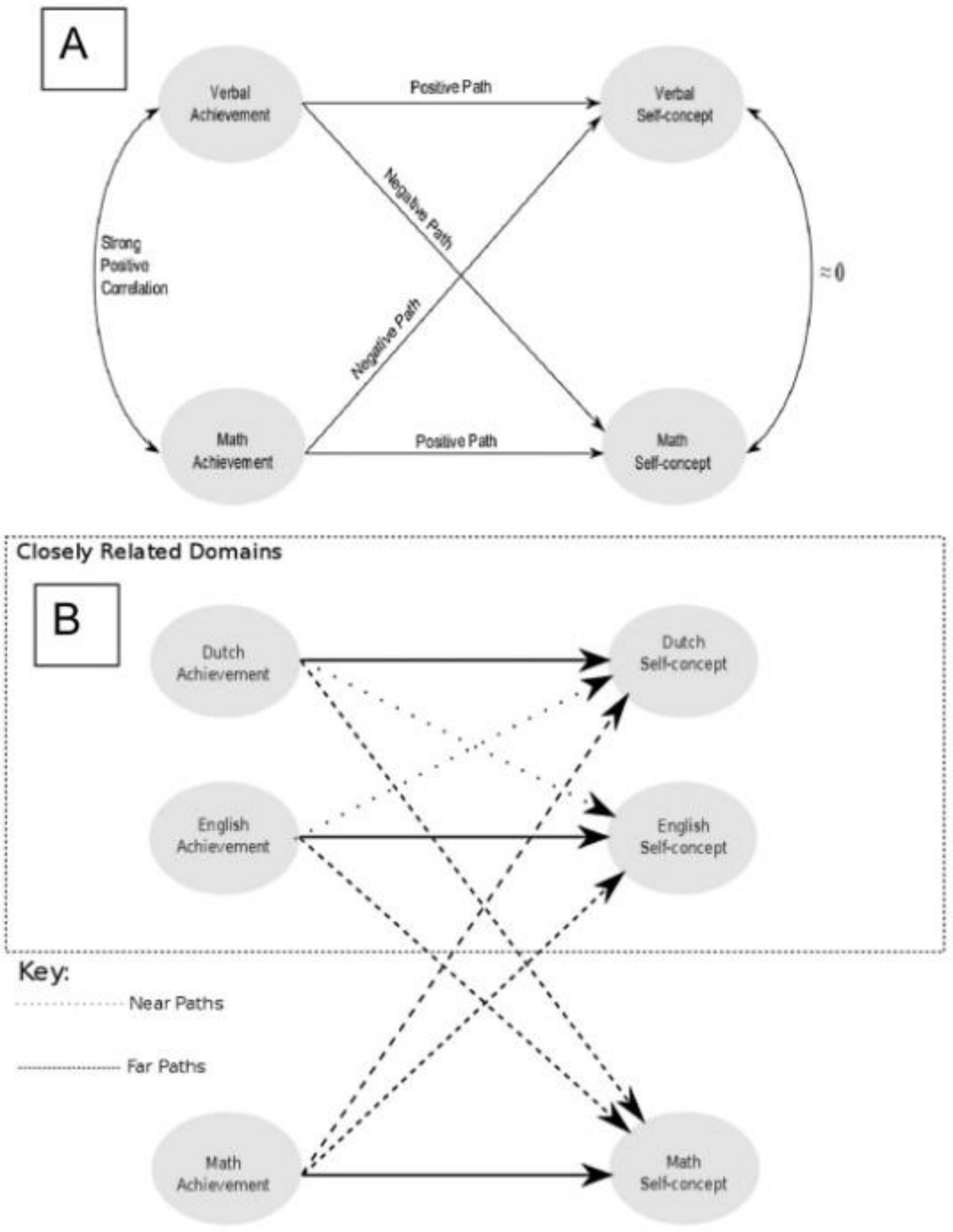

Fig. 1. A: The ‘Classic’ Internal/External Frame of Reference (I/E) Model relating verbal and math achievement to verbal and math self-concept. According to predictions from the I/E model, the horizontal paths from achievement to self-concept in the matching domains (content area) are predicted to be substantial and positive, whereas the cross paths from achievement in one domain area to self-concept in a non-matching domain are predicted to be negative (as contrast). B: Extending I/E model to include two closely related verbal domains. Far cross paths (relating math and the two verbal domains) are again predicted to be negative. However, the near cross paths (relating the two verbal domains) are predicted to be significantly less negative, non-significant, or even positive (assimilation).

In a large cross-cultural study, Marsh and colleagues (Marsh, Hau, 2004, Marsh et al, 2006) demonstrated that support for these theoretical predictions generalized over large, nationally representative samples of 15-year-olds from 26 countries. In a meta-analysis of 69 data sets Möller, Pohlmann, et al. (2009) reported that math and verbal achievements were highly correlated (.67), but self-concepts were nearly uncorrelated (.10). The horizontal paths from achievement to ASC in the matching domains were positive (.61 for math, .49 for verbal), but cross paths were negative from math achievement to VSC (−.21) and verbal achievement to MSC (−.27). Strong support for the generalizability of the I/E predictions led these authors to conclude, “The results of our meta-analyses indicate that the relations described in the classical I/E model are not restricted to a particular achievement or self-concept measure or to specific age groups, gender groups, or countries” (p. 1157), making it one of the most robust empirical findings in educational psychology research.
Providing stronger tests of causal mechanisms posited in the theoretical I/E model, Möller and colleagues (e.g., Möller, 2005, Möller, Köller, 2001a, Möller, Köller, 2001b, Möller, Savyon, 2003, Pohlmann, Möller, 2009) experimentally manipulated the external (social) comparison process based on performance feedback relative to other students, and the dimensional comparison process based on feedback relative to performances by the same student on two subject-specific tasks. These true experimental studies provided strong support for causal interpretations of both the dimensional and the social comparison processes posited in the I/E model. In two introspective diary studies, Möller and Husemann (2006)also confirmed that students spontaneously carry out dimensional comparisons in everyday life, with negative (contrast) effects from one domain to self-evaluations and emotions in the other.

The I/E model has also been heuristic in relation to other major theoretical models in educational psychology. For example, Pekrun (2006; see also Goetz et al, 2010, Goetz et al, 2008, Goetz et al, 2006) has demonstrated that theoretical predictions based on the I/E for self-concept responses also generalize to emotional responses, and has incorporated the I/E model into his control-value theory of achievement emotions. Similarly, Eccles and colleagues (Eccles et al, 2004, Nagy et al, 2008, Parker et al, 2012, Parker et al, 2012) integrated support for I/E predictions into expectancy-value theory for the prediction of gender differences in academic and career choice.

2.2. Dimensional comparison theory (DCT)

The current investigation is the first empirical test of a recently published theoretical account of DCT (Möller & Marsh, 2013; see also Marsh et al., 2013) and places the I/E model in a much broader, more general framework. Here we focus on three new theoretical contributions.

2.2.1. Near vs. far comparisons

Empirical support for the classic I/E model (Fig. 1A) is based mainly on math and verbal domains, so that there are only “horizontal” paths between matching domains and “cross” paths between non-matching domains. Although several I/E studies have considered additional domains (e.g., Bong, 1998, Marsh et al, 2001, Marsh, Yeung, 2001, Möller et al, 2006, Xu et al, 2013), none of these was based on the new theoretical DCT framework, which incorporates domains other than the math and verbal domains emphasized in the classic I/E model (Fig. 1B). Hence, a critically important theoretical contribution of DCT is its expansion of the scope of the theoretical perspective to include a wider variety of domains, in which the cross-paths involve “near” and “far” comparisons in relation to how similar or dissimilar different school subjects are to each other. This ordering of school subjects along an a priori verbal-to-math continuum is based on theoretical and empirical research that led to the Marsh/Shavelson revision (Marsh, 1990, Marsh et al, 1988, Marsh, Shavelson, 1985) of the original Shavelson et al. (1976) multidimensional, hierarchical model of self-concept (see Marsh, 2007), thus integrating DCT with established self-concept theory and empirical results. In this theoretical model of the structure of ASC, the ordering of different domains is based on relations among the domains derived from a higher-order factor analysis of the multiple ASC domains (Fig. 2). Of particular relevance to the present investigation, there is a substantial literature showing that ASCs in native and foreign languages are more closely related to each other than to math self-concept (e.g., Xu et al., 2013).

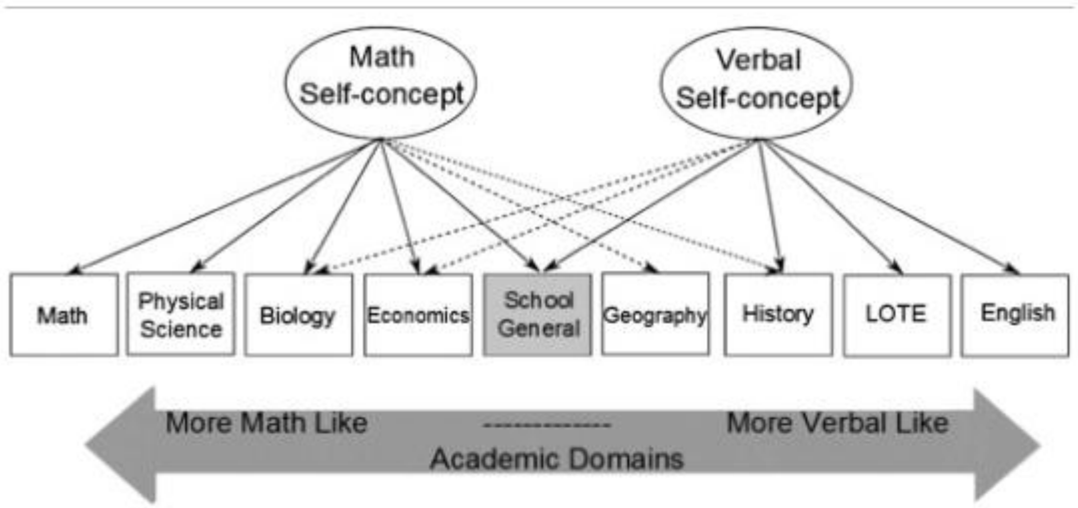

Fig. 2. The Marsh/Shavelson revised theoretical model of the structure of academic self-concept. Adapted with permission from Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in Educational Psychology*. Leicester, UK: British Psychological Society.

Particularly for social comparison studies in the self-concept literature (see review by Marsh, 2007, Marsh et al, 2008), the predominant finding is one of contrast; if classmates are more able, then a student's self-concept will be lower. In the broader social psychology literature, covering a diversity of domains—mostly not academic in nature—there is also a predominance of contrast effects (e.g., Diener & Fujita, 1997). Even though some theoretical models in the social psychology literature suggest assimilation effects (e.g., Brickman, Bulman, 1977, Collins, 1996, Tesser, 1988; sometimes referred to as positive contextual effects, reflected glory effects, or contagion effects), replicable support for assimilation effects continues to be elusive (Diener, Fujita, 1997, Marsh et al, 2008, Marsh et al, 2010). Hence, a critical feature of the expansion of the I/E model into DCT is to highlight this theoretical distinction between near and far comparisons, testing the a priori prediction that contrast effects based on “near” comparisons will be substantially less negative and might even be positive rather than negative (i.e., assimilation effects). Hence, similar subjects (e.g., native and foreign languages) might be seen as complementary, rather than contrastive, such that achievement in one domain contributes positively to self-concept in a complementary, near domain. This is also important in building bridges between research based on the classic I/E model studied in educational settings, and the more diverse social comparison literature. Based upon our integration of this theoretical material into DCT, in the present investigation we hypothesize that:

- Cross paths relating “far” domains (math with English and Dutch) will be significantly negative (contrast); and
- Cross paths relating “near” domains (English and Dutch) will be significantly less negative than cross paths relating far domains, and may be non-significant or even positive (an assimilation).

2.2.2. Achievement: class marks vs. standardized test scores

The present investigation is apparently the first to integrate into DCT the findings of previous self-concept research and new theoretical predictions about the distinct predictions of ASC based on class marks (i.e., school grades on report cards) and standardized test scores. Class marks provided by classroom teachers typically are the most immediate and salient source of feedback that students receive about their academic accomplishments, and thus have high ecological validity. However, class marks have idiosyncrasies that complicate their interpretation. For example, teachers tend to grade on a curve, such that the best and worst students in each class tend to get the highest and lowest marks, independently of the average ability levels of students within each class. Furthermore, grading standards and the marking basis of class marks (e.g., various combinations of standardized tests, classroom tests, homework, projects, good behavior, and classroom participation) also tend to be somewhat idiosyncratic to particular subjects and individual teachers. Standardized test scores have the advantage of providing a common metric to evaluate the achievement of all students, independently of their class, teacher, or school. However, typically this is a less immediate, less salient source of feedback to students—particularly for low-stakes tests in which students might not even receive feedback about their performance. Although previous self-concept research has emphasized the importance of this distinction in relation to the size of correlations based on matching areas of achievement and self-concept, here we build upon it by integrating it more fully into DCT.

Early self-concept studies (e.g., Hansford, Hattie, 1982, Marsh, 1987; see discussion by Marsh, 2007, Marsh, Craven, 2006) provided a theoretical rationale and empirical support for the expectation that academic self-concept should be more related to achievement in the matching domain when based on class marks than when based on standardized achievement tests. This rationale is also consistent with theoretical perspectives from social psychology research on the local dominance effect (Alicke et al, 2010, Zell, Alicke, 2009), in which the most local frame of reference is the most important determinant of self-evaluations. In these social comparison studies of the local dominance effect, in which “local” and more “general” frames of reference are experimentally manipulated, participants consistently used the most local comparison information available to them, even when they were told that the local comparison was not representative of the broader population and they were provided with more appropriate normative comparison data. Because class marks are more local than test scores, the local dominance effect predicts that class marks should be more related to ASCs. Thus, the rationale from social psychology is complementary to the rationale developed within self-concept research in educational psychology.

Consistent with previous self-concept research, in their meta-analysis of I/E studies, Möller et al. (2009) found that the largest moderator of I/E results was the nature of the achievement measure. More specifically, the correlation between math and verbal achievement was substantially larger when based on standardized test scores than on class marks, whereas academic self-concepts were more strongly correlated with achievement in the matching domain when based on teacher-assigned marks rather than standardized tests. However, the meta-analysis results were inherently weak, in that conclusions were necessarily based on different studies that variously considered either class marks or test scores, rather than specifically comparing results for the two indicators of achievement based on the same students in a single, large-scale study. Extending the logic of DCT, we predict that cross-paths based on far comparisons should be more negative when based on class marks than when based on standardized test scores. Although this is not specifically evaluated in the Möller et al. (2009) meta-analysis, their results apparently do not support this prediction. However, we again note that their meta-analysis—based largely on studies either of test scores or class marks—is not particularly strong in terms of evaluating differences between test scores and class marks, so that it is relevant to explore further support for this theoretical prediction.

What of cross-paths based on near comparisons? Because there is no clear evidence even for the direction of these paths (i.e., assimilation or contrast), there is even less basis for predicting how these paths are likely to differ for class marks and test scores. Indeed, there is no previous empirical evidence on this issue and our study is apparently the first even to pose this issue as a theoretical question. Hence, we leave as a research question whether the size or direction of near comparisons differs for test scores and class marks; this is an issue that has not previously been studied or even speculated upon.

Pre-empting further discussion, we also note that the present investigation is ideally suited to testing this distinction between standardized test scores and class marks, due to unique features of the Dutch secondary school system (see subsequent discussion; see also Meelissen, Punter, 2012, Scheerens et al, 2011, Centre on International Education Benchmarking,. Our study is based on a large, representative sample of Dutch secondary schools (15,356 9th grade high school students from 651 classes in 95 schools). In the Dutch school system, classes within schools are highly tracked. Students in the high-track classes perform substantially better than students in lower-track classes, when evaluated by achievement measured along a common metric (i.e., standardized tests). Furthermore, the composition of each class is the same for different school subjects, so that students are only exposed to other students of similar abilities who are in the same track. However, class marks in these highly tracked classes depend substantially on how students compare with other students in the same class, whereas standardized tests rank all students in relation to a common metric that is independent of the class, track and school. Hence, the distinction between class marks and standardized test scores is stronger and more clearly defined in the Dutch system than is typical in previous research on this topic. Following from our integration of this previous research into DCT, we hypothesize that:

- ASCs will be more strongly correlated with matching class marks than with test scores;

- Correlations among the different test scores will be substantially larger than those among the class marks, and correlations among the class marks will be substantially larger than those among the ASCs (see Möller et al., 2009, meta-analysis);
- Horizontal paths in the I/E model (Fig. 1) will be significantly positive for both class marks and test scores considered separately. However, when both sets of achievement measures are included in the same model, horizontal paths will be more positive for class marks than for test scores; and
- Cross paths based on far comparisons (Fig. 1B) will be significantly negative for both class marks and test scores considered separately, but there will be

more negative cross paths for class marks than test scores when both sets of achievement measures are included in the same model. We leave as a research question how near paths will differ when based on class marks and test scores, as there is no theoretical basis for making a priori predictions, nor any prior research.

2.2.3. MeVclass ratings: distinguishing between social comparison and dimensional comparison processes

Early criticisms of the I/E model (e.g., Bong, 1998, Dai, Rinn, 2008; but see responses by Marsh, Yeung, 2001, Marsh et al, 2008) questioned the validity of the interpretation that the negative contrast effects (i.e., the negative cross-paths in Fig. 1A) are actually due to dimensional comparison processes. However, subsequent research, particularly by Möller and colleagues (e.g., Möller, 2005, Möller, Köller, 2001a, Möller, Köller, 2001b, Möller, Savyon, 2003, Pohlmann, Möller, 2009), provides much stronger support for the theoretical rationale underlying these predictions in relation to true experimental studies, in which the students are randomly assigned to conditions in which the frame of reference is experimentally manipulated. Nevertheless, in relation to DCT predictions, it continues to be difficult to separate the predictive effects of social and dimensional comparison processes.

In a theoretical advance in the social psychology research literature, Huguet et al. (2009) proposed global meVclass ratings (ratings by each student as to how they compare to other students in the same class: 1 = *much worse* to 5 = *much better*) as a pure measure of social comparison processes. They provided strong empirical support for their theoretical tenet that meVclass ratings are a pure indicator of social comparison processes, for the construct validity of meVclass ratings, and for the claim that their research was the first empirical demonstration that negative contextual effects associated with class-average ability were actually due to social comparison processes. In the present investigation, we extend this work by incorporating meVclass ratings in specific school subjects and integrating them into DCT, to provide new tests of the distinction between social and dimensional comparison processes. In particular, we argue that by controlling for meVclass ratings, we are controlling for all or most of the social comparison effects, thereby isolating the dimensional comparison effects and substantially unconfounding the otherwise confounded effect of social and dimensional comparison effects. Based on this integration of meVclass ratings into DCT, we hypothesize that:

- Correlations among meVclass ratings in specific subjects (Dutch, English, math) will be higher than correlations among the corresponding ASCs, but smaller than correlations among the corresponding test scores and similar in size to correlations among class marks. This follows from our assumption that meVclass ratings are primarily a function of class marks and are relatively free of internal dimensional comparison processes, and from previous research (e.g., the Möller et al., 2009 meta-analysis) showing that correlations among test scores in different domains are systematically higher than correlations among the corresponding class marks;
- Consistent with the multidimensional perspective and support for discriminant validity of meVclass ratings, the global meVclass ratings are predicted to be substantially less correlated with domain-specific measures of achievement and ASC than the domain-specific meVclass rating in the matching domain.

- When meVclass ratings are added to the I/E model, the positive horizontal paths from achievement to matching ASCs will be substantially smaller, as the horizontal paths substantially reflect the social comparison process represented by the meVclass ratings. However, the negative cross paths from achievement to a non-matching domain will be much less affected, if affected at all, as these are hypothesized to reflect primarily dimensional comparison processes that are relatively unrelated to the meVclass ratings.

3. Method

3.1. Participants, measures, and procedures

3.1.1. Participants

Our study is based on a large, nationally representative sample of 15,356 Dutch 9th grade high school students from 651 classes in 95 schools. Students were in their third year of Dutch secondary education (corresponding to US grade 9; Mn age = 14.7 years, SD = 0.7). The sample was balanced in relation to gender (51% girls).

Dutch secondary education is one of the most highly tracked school systems in the world (for a more detailed description of the Dutch school system, see Meelissen, Punter, 2012, Scheerens et al, 2011, Centre on International Education Benchmarking,). The Dutch secondary system has five different tracks, which differ in difficulty level, duration, content, and admission rights for further education. The lowest three tracks are classed together as pre-vocational education (basic track, middle track, theoretical track) and include four years of schooling, after which these students typically follow senior vocational education. The next track is senior general secondary education and includes five years of schooling, which qualifies students for higher professional education. The highest track is pre-university education, which takes six years, after which students qualify for university. Placement in the first year of secondary education is determined largely by the score on a standardized test taken at the end of primary school (students aged 11 or 12), based on language, arithmetic/mathematics, and learning skills, and by recommendations from primary school teachers. Given the test score and recommendations, the children and their parents select a secondary school. Although many schools have mixed-ability classes in the first year of high school, by the third year (the basis of the present investigation) almost all classes contain students from a single track. Moreover, the class—the group of students sitting in the same classroom—is the same for different school subjects. Because student class composition is the same for all school subjects, it greatly facilitates evaluations of contextual predictive effects, as the composition of classmates (and thus, the frame of reference) does not vary for different school subjects (e.g., Seaton et al., 2008).

3.1.2. Measures

The standardized achievement tests were developed and scored by CITO—a Dutch institute for testing services [see: http://www.cito.com/]. The three tests (Dutch reading comprehension, English reading comprehension, and mathematics) each had three versions, designed for students from different tracks, with a partial overlap to enable equating of tests. The items were scored with the one-parameter-logistic-model, resulting in unique ability estimates (theta-scores) for each student (Hambleton, Swaminathan, & Rogers, 1991). Class marks were based on report card results from the previous semester, assigned by the different teachers in each of the three subjects (Dutch, English, mathematics). These were based on class marks provided by teachers on the last report card prior to the completion of the survey. As part of the survey, ASC in each of the three subjects (Dutch, mathematics, and English) was determined by responses to three questions: “How good are you in school in [Dutch, mathematics, and English]” (1 = *not good at all* to 7 = *very good*); “How easy or difficult do you find [Dutch, mathematics, and English]” (1 = *very difficult* to 7 = *very easy*); “How quickly do you learn new things in [Dutch, mathematics, and English]” (1 = *very slowly* to 7 = *very quickly*). Based on data in the present investigation, the coefficient alpha estimates of reliability for each of these three item scales were: .77, Dutch; .86, mathematics, .83, English. We note, however, that measurement error was automatically controlled in the latent variable models considered here. Strong support for the a priori factor structure underlying responses to these ASC items is presented in the external Supplemental Materials (available on this Journal's website).

Students also completed meVclass ratings, comparing themselves to the other students in the class in each of the three school subjects and in global achievement: “Do you think that you, in comparison with most of your classmates, are better or worse in [Dutch, mathematics, English, global achievement]” (1 = *much worse* to 5 = *much better*).

3.1.3. Procedures

Data used in the present investigation are part of the larger Cohort Onderzoek Onderwijs Loopbanen (COOL), a cohort study of Dutch secondary education (for a description of the Dutch school system, see Meelissen, Punter, 2012, Scheerens et al, 2011, Centre on International Education Benchmarking,; see earlier discussion). These data were collected during the first half of 2011. All secondary schools were informed about the COOL project and the probability that they would be asked to participate. The initial sample of secondary schools was designed to maximize inclusion of the number of students who had previously been involved in a related study, made when these students were in primary schools, but also their new classmates. Some schools only agreed to collect data from students who had been in the previous study, or for a sample of students. However, in many schools—those considered in the present investigation—all students who were in attendance on the day of data collections, in all classes in the year group, participated, although a small number of individual students or their parents had refused participation. There is also a missing-by-design issue in that, while at all schools, all students completed Dutch and mathematics tests, schools could choose to take English language or civics or both. However, class marks, meVclass ratings, and ASC responses were collected for all students in Dutch, mathematics, and English.

3.2. Data analysis

All analyses were conducted with Mplus 7.11 (Muthén & Muthén, 2008–2013). Confirmatory factor analyses (CFAs) and structural equation models (SEMs) used the robust maximum likelihood estimator (MLR), with standard errors and tests of fit that were robust in relation to non-normality of observations and the use of Likert responses (e.g., Beauducel, Herzberg, 2006, DiStefano, 2002, Dolan, 1994, Muthén, Kaplan, 1985). MLR estimation is also robust to the non-independence of observations when used in conjunction with a design-based correction (Mplus's complex design option; Muthén & Muthén, 2008–2013) that controls for the hierarchical, nested nature of the data. The amounts of missing data varied depending on the variable: school, class and track identification (0%), self-concept responses (11–12%), meVclass ratings (17%), test scores (15% Math; 18% Dutch; 37% English), gender (1%), age (2%). We used full information maximum likelihood (FIML) estimation to control for missing data to obtain unbiased parameter estimates, standard errors, and goodness of fit statistics.

In preliminary analyses (presented more fully in the Supplemental Materials), we evaluated the factor structure of the ASC responses. Because each factor was based on parallel worded items, we posited correlated uniquenesses relating each item with the same wording, noting that the failure to do so would substantially diminish fit to the data and result in biased parameter estimates (see Jöreskog, 1979, Marsh, Hau, 1996). Consistent with expectations, the model with correlated uniquenesses fitted much better (goodness of fit statistics and parameter estimates are presented in Table 1 of the Supplemental Materials). In the next model, we added the set of six achievement indicators (standardized test scores and class marks in Dutch, English, and math) and the set of four social comparison (meVclass) ratings (see Supplemental Materials). We examine the correlations among these factors in the results section (goodness of fit statistics and parameter estimates are presented in the Supplemental Materials, along with further discussion of the results).

In an apparently new and potentially important contribution to DCT research, we used Mplus's model constraint procedure (Muthén & Muthén, 2008–2013), based on the delta method (Oehlert, 1992) to estimate the average parameter across different domains and to formally test the statistical significance of differences hypothesized earlier (e.g., differences between correlations based on test scores and class marks; differences between horizontal and cross paths; differences between paths based on test scores and school marks; and differences between far and near cross paths). The critical advantage of this procedure is that the average values are based on results of appropriate latent variable models; model-based, more appropriate standard errors for confidence intervals and hypothesis tests can then be obtained by applying the multivariate delta method (see Raykov & Marcoulides, 2004, for an accessible introduction).

4. Results

4.1. Relations among constructs: construct validity of ASCs and social comparison ratings

4.1.1. ASC factors: convergent and discriminant validity

We begin with an evaluation of relations among constructs (Table 1) to evaluate support of the convergent and discriminant validity of the ASC responses, but also as an advanced organizer for subsequent analyses. The average correlation among the three latent ASC factors is close to zero (Mn $r = -.013$, SE = .008; see Table 2). Although there is a small positive correlation ($r = .190$) between Dutch self-concept (DSC) and English self-concept (ESC), both these verbal ASCs are negatively correlated with MSC (rs = –.060 and –.169, respectively). Thus, the ASC factors are remarkably distinct, in comparison to correlations among the corresponding class marks (Mn $r = .259$, SE = .010) and particularly the test scores (Mn $r = .539$), which were all positive. Particularly the relations among the three ASC domains support theoretical predictions about the ordering of domains posited in the theoretical structure of ASC (see Fig. 2) that is the basis for the theoretical distinction between near and far comparison in DCT.

In support of convergent validity, ASCs are significantly correlated with the matching achievement based both on class marks (Mn $r = .693$, SE = .006) and on test scores (Mn $r = .296$, SE = .008). However, consistent with a priori predictions, correlations with class marks are much higher than those based on test scores (difference = .693 – .296 = .397, SE = .009; Table 2). This pattern of results, the extreme domain specificity of ASCs in different school subjects in relation to each other and to measures of achievement, and the substantially higher correlations with class marks than with test scores, is consistent with previous research and with a priori predictions (e.g., Marsh, 2007; also see earlier discussion). The results provide strong support for both the convergent and discriminant validity of the ASC responses. It is also interesting to note that support for the domain specificity of the class marks, and particularly for the test scores in different school subjects, is much weaker than for ASCs. Thus, the mean correlation among the three domains for test scores is substantially higher than for the ASCs (Table 2).

4.1.2. Near vs. far domains

Based on the academic continuum of ASCs (Fig. 2), we classified Dutch and English as “near” domains, while math was classified as a “far” domain in relation to each of these two verbal domains (see Fig. 1B). Support for a priori predictions and for this classification in relation to ASCs is clearly evident; the correlation between DSC and ESC is significantly positive (.190, Table 1), while correlations relating MSC to DSC and ESC are both negative (–.06 and –.169, respectively). Although this academic continuum is posited in relation to ASCs rather than to achievement measures, it is interesting to note that the same pattern of differences is also evident in correlations among the class marks, even though all the correlations are positive. However, correlations among the three test scores are all similar (.519 to .574), and not even the direction of the small differences is consistent with our near and far distinction. Thus, consistent with a priori predictions, a general factor dominates test scores more than is the case for class marks, and particularly with ASCs.

4.1.3. MeVclass ratings

A unique feature of the present investigation is the addition of meVclass ratings, which are designed to measure social comparison processes. They are important in testing processes underlying the I/E model, but are also of interest in their own right. Consistent with a priori predictions and the discriminant validity of meVclass ratings, the global meVclass ratings are substantially less correlated with achievement and ASC measures than with the domain-specific meVclass rating in the matching domain. Indeed, the domain specificity—both the size and pattern—of the meVclass ratings is surprisingly similar to those observed for the ASC factors (Mn $r = .030$, SE = .007). Furthermore, the correlations between these single-item meVclass ratings and matching latent ASC factors are very high ($r = .706$ to .854, Table 1; Mn $r = .795$). There is also support for the convergent and discriminant validity of the meVclass ratings in relation to test scores (Mn $r = .272$, SE = .010), and particularly to school marks (Mn $r = .587$, SE = .006). However, the extreme domain specificity of the meVclass ratings also leads to the rejection of our prediction that intercorrelations among the meVclass ratings would be substantially greater than intercorrelations among ASC factors, and more similar to intercorrelations among the corresponding test scores. Although correlations among meVclass ratings are significantly larger than those among the ASC factors (Mn $r = -.043$, SE = .007), the differences are small, and they are substantially less than intercorrelations among class marks (Mn $r = .229$, SE =.010).

Although this is tangential to the present investigation, responses to meVclass ratings typically result in what has been referred to as the better-than-average predictive effect (e.g., Kuyper, Dijkstra, Buunk, & Van der Werf, 2011), in which the average rating across all respondents is substantially higher than the “average” response category. However, for our data, mean meVclass ratings were all close to 3.0, the mid-point of the response scale, suggesting little systematic response bias (Dutch, 2.9; English, 3.1; mathematics, 3.0; global school, 3.1). These mean ratings, coupled with the very high correlations between classVme ratings and corresponding class marks (as well as ASCs) support the convergent and discriminant validity of these responses.

In summary, there is good support for the domain specificity, and for the convergent and discriminant validity of the meVclass ratings. However, these results also call into question the interpretation of the meVclass ratings as pure measures of social comparison, and contradict a priori predictions that correlations among these ratings would be substantially higher than correlations among the corresponding ASC factors. Indeed, the pattern of results suggests that meVclass ratings behave more like ASC ratings than do class marks or test scores, suggesting that internal dimensional comparison processes might also be influencing these meVclass ratings. We explore this issue further in the following tests of the I/E model and its extension.

4.2. Tests of the I/E model and its extension

4.2.1. Initial evaluation of traditional I/E predictions

In Table 3A, Table 3B we present results from five alternative models designed to test a priori predictions based on the I/E model and its extension. Models 1–5 are SEMs in which the three latent ASCs (DSC, ESC, and MSC) are regressed on the set of three test scores (Model 1), the set of three class marks (Model 2), the four meVclass ratings (Model 3), the combined set of test scores and class marks (Model 4), and the combined set of test scores, class marks, and meVclass ratings (Model 5). In traditional I/E models, the focus is on verbal (native-language) and math constructs. In all five models summarized in Table 3A, Table 3B, there is clear support for the traditional I/E model (Fig. 1A) based on math and verbal (Dutch) constructs. In particular, the horizontal paths relating predictors to matching ASCs (i.e., math predictors to MSC, Dutch predictors to DSC) are substantially positive, while the cross-paths relating predictors to non-matching ASCs (i.e., math predictors to DSC, Dutch predictors to MSC) are significantly negative. Although they are clearly supportive of the traditional I/E predictions, these results are not surprising, given the substantial body of support for these predictions. Hence, we now pursue more detailed tests based on our main research hypotheses.

4.2.2. Horizontal paths: test scores vs. class marks

Horizontal paths (see Fig. 1) are those leading from achievement in each subject to the corresponding ASC factor. Consistent with predictions, all of these are substantial and positive. However, also consistent with predictions, these standardized path coefficients are substantially higher for class marks (Mn = .759, SE = .008; Model 2, Table 3A, Table 3B) than for test scores (Mn = .408, SE = .408, SE = .011; Model 1, Table 3A, Table 3B).
In Model 4 we directly pitted these two sets of predictors (test scores vs. class marks) against each other by including both in the same model. Not surprisingly, the unique contribution of each was diminished. However, the horizontal paths based on class marks in Model 4 were nearly as strong as those in Model 2 (Mns = .715 vs. .759), while those based on test scores were substantially weaker (Mns = .177 vs. .408). Consistent with a priori predictions, the difference between horizontal paths based on test scores and class marks (.715 – .177 = .538, SE = .012) is statistically significant.

4.2.3. Cross paths: near vs. far domains

The most striking feature of the I/E model is the negative cross paths relating achievement in one domain to ASC in another domain (contrast). However, based on the original theoretical foundation of the I/E model and its extension into DCT, the sizes of these cross-paths should vary substantially, depending on the nature of the two domains. Indeed, it is reasonable to expect that higher achievements in one domain might even have a positive (assimilation) predictive effect on ASCs in a closely related (near) domain. A priori we classified DSC and ESC as near in relation to the ASC continuum (see Fig. 2 and empirical support based on correlations from this study discussed earlier), while MSC was classified as far, in relation to these two language domains.

For test scores (Model 1, Table 3A, Table 3B), the mean cross paths averaged across near and far domains is significantly negative (Mn = –.101, SE = .006). The mean cross paths for the far domains (math vs. Dutch and English) are negative (Mn = –.182, SE = .008). However, the mean cross path for near domains (Dutch and English) is significantly positive (Mn = .061, SE = .011); an assimilative prediction effect. Although only one of these two paths is statistically significant when considered separately (the path from English tests to DSC is non-significant), even the non-significant correlation is consistent with a priori predictions that near paths are less negative (or more positive) than far paths. Hence, consistent with a priori predictions, the difference between these far and near paths (–.182 – .061 = –.243, SE = .015) is significant.

For class marks (Model 2, Table 3A, Table 3B), the average cross-path is negative (–.140, SE = .004). Here, however, the mean near (–.086) and far (–.167) paths are both significantly negative (contrast). Nevertheless, consistent with predictions, the far paths are significantly more negative than the near paths (difference = .081, SE = .009).
When we combine test scores and class marks in the same model (Model 4, Table 3A, Table 3B), the pattern of results is similar. For class marks, both far paths (Mn = –.151, SE = .005) and near paths (Mn = –.092, SE = .008) are significantly negative, but far paths are significantly more negative (difference = .059, SE = .010). For test scores, far paths are significantly negative (Mn = –.070, SE = .006), while the average near path is not significantly different from zero (.010, SE = .009). However, far paths are again significantly more negative (diff = –.080, SE = .012). Consistent with a priori predictions, the results in Model 4 show that the cross paths based on class marks are significantly more negative than those based on test scores for both near (.102, SE = .014) and far (.081, SE = .009) paths.

4.2.4. MeVclass ratings and social comparison

Incorporation of meVclass ratings represents an apparently new addition to DCT studies. Following Huguet et al. (2009) we posited these to be direct measures of social comparison, thus providing a basis for the disentangling of the dimensional and social comparison processes underlying the I/E model. Consistent with these expectations (and the results from the social comparison study of Huguet et al.), we predicted that including meVclass ratings would substantially reduce the sizes of horizontal paths but would not substantially affect the cross-paths. However, our preliminary evaluation of correlations among the various constructs (see earlier discussion of Table 1) suggests that meVclass ratings might be influenced by both social and dimensional comparison processes in much the same way as ASC responses. Here we explore issues in a number of different models.
In Model 3, we simply treat the meVclass ratings as a separate set of predictors of ASC factors, albeit based on students' subjective perceptions of how they compare with other students within their class in each of the three subjects, and globally. In this respect, it is

important to note that all students in this study have the same classmates for each subject (i.e., class composition is the same for different subjects). In this model, the horizontal paths are very large (Mn = .781, SE = .009), far cross-paths are significantly negative (Mn = –.072, SE .006), and the mean of the near cross-paths is not statistically significant (Mn = .000, SE = .006). However, the far paths are significantly more negative than the near paths (difference = .072, SE = .006). For meVclass ratings, there was also a global academic component, with small positive paths to each ASC (Mn = .044, SE = .005); much smaller than the horizontal paths but more positive than the cross-paths. Hence, at least the pattern of results is consistent with expectations; horizontal paths based on meVclass ratings are higher than those based on class marks and particularly those on test scores.

In Model 5, we included test scores and class marks (as in Model 4) and meVclass ratings (as in Model 3). The inclusion of meVclass ratings substantially reduced but clearly did not eliminate the positive horizontal paths from test scores (Mn = .090, SE = .006), and particularly not those from class marks (Mn = .385, SE = .006), or the difference between the two (.385 – .090 = .295, SE = .010). Not surprisingly, however, the largest horizontal paths in Model 5 were from the meVclass ratings (Mn = .544, SE = .007).

Cross paths in Model 5 were also substantially reduced, but not eliminated, by the inclusion of meVclass ratings. For test scores the near cross paths were small, but significantly positive (Mn = .015, SE = .007), while those for the far cross paths were small but significantly negative (Mn = –.036, SE = .005). For class mark scores, both the near (Mn = –.076, SE = .008) and far (Mn = –.092, SE = .006) paths were significantly negative.

Of particular interest in relation to a priori predictions is the comparison of Model 4 (without meVclass ratings) and Model 5 (with meVclass ratings). Consistent with predictions, the inclusion of the meVclass ratings substantially diminished the horizontal paths from test scores (.177 – .090 = .087), particularly with class marks (.715 – .385 = .330). The inclusion of meVclass ratings also diminished the negative cross paths, but to a much smaller degree for both test scores [–.043 – (–.019) = –.024]) and (particularly) class marks [–.131 – (–.087) = .044]. Interestingly, the near cross paths (.032, SE = .006) for the meVclass ratings were significantly positive, while the far cross paths were not statistically significant (–.006, SE = .005). The difference between the two was statistically significant (.032 + .006 = .038, SE = .008). Hence, for the meVclass ratings in Model 5, there were no significant negative cross paths.

In summary, there was partial support for the predictions based on meVclass ratings. Consistent with predictions, the inclusion of meVclass ratings substantially reduced the size of the positive horizontal paths. Also consistent with predictions, these reductions in the horizontal paths were substantially larger than the corresponding reductions in the cross paths. Nevertheless, contrary to the predictions noted earlier, inclusion of the meVclass ratings results in a small but statistically significant change in the cross paths. Also contrary to predictions, correlations among the meVclass ratings (see earlier discussion of Table 1, Table 2) were almost as small as correlations among the ASC factors, and much smaller than those among the class marks.

4.2.5. Supplemental models of meVclass ratings and social comparison

In order to explore further the role of frame of reference effects in the formation of meVclass ratings, we tested several supplemental models. In Models 6A–6C (Table 4A, Table 4B), we treated the meVclass ratings (instead of the ASC factors) as the outcome variable predicted by the test scores and class marks. Thus, Models 6A–6C, with meVclass ratings as the outcome variable, parallel Models 1, 2, and 4, with ASC as the outcome variable. The pattern of results for these models largely parallels those based on ASCs, in that the horizontal paths are positive, the far cross paths are negative, and the near cross paths are small and sometimes even positive, rather than negative (see Table 4A, Table 4B). The mean paths are mostly smaller in Model 6C, but the pattern of results (direction and statistical significance) is very similar. Of particular relevance, the mean cross path is significantly negative for both test scores and class marks, although the mean far cross path is significantly more negative than the mean near cross path. More clearly than earlier results based on the meVclass ratings, these results suggest that there is a dimensional comparison process in the formation of meVclass ratings.

In Models 7 and 8 we fitted models relating test scores and class marks. In Model 7 class marks were predicted on the basis of test scores. Here there is a clear pattern of I/E predictive effects in which horizontal paths are positive (Mn = .305, SE =.010), near paths are less positive (Mn = .089, SE =.013), and far paths are significantly negative in each of the three domains (Mn = –.084, SE = .008). In particular, the significantly negative far paths suggest a process like dimensional comparison. However, in Model 8, in which test scores were predicted by class marks, no cross paths were significantly negative. The juxtaposition of these two models suggests that there is a process akin to the dimensional comparison process in the formation of class marks that is not evident in test scores. In interpreting these results, it is important to emphasize that the teachers for each subject were different, so that the process is unlikely to be the result of strategies used by individual teachers in assigning class marks. Rather, we interpret the results to be a function of student strategies. Thus, we conjecture that students are more motivated, more conscientious, and put in more effort in those subjects in which they feel more competent. At least in the short term, these psychological processes are likely to have more influence on class marks assigned by teachers on the basis of classroom performance, than on low-stakes standardized tests that have no direct implications for the students. This dimensional comparison predictive effect observed in the class marks also explains in part why there is a dimensional comparison predictive effect in the meVclass ratings.

4.2.6. Supplemental models controlling for background/demographic variables

Model 4 (Table 3A, Table 3B) provides clear support for DCT predictions, particularly in relation to differences between the near and far paths that is the distinctive new feature of DCT in relation to historical support for the I/E model. Here we briefly evaluate how the effects change when additional background/demographic variables are included as covariates (Model 4A in Table 5): academic track, gender, age, SES, and ethnicity (Western vs. non-Western). The only coefficients to change in terms of statistical significance or direction were the two near paths from English Test to Dutch Self-concept (−.001 SE = .015 in Model 4 to .032 SE = .016 in Model 4A, Table 5) and Dutch Test to English Self-concept (.021 SE = .011 in Model 4 to .090 SE = .011 in Model 4A, Table 3A, Table 3B). Hence, in both cases these small differences associated with the inclusion of these covariates are in the direction of a priori predictions. Other changes are that the positive effects of test scores on ASCs in the matching area became more positive. In summary, inclusion of additional covariates had little effect on support for DCT and led to marginally stronger support for the distinction between near and far paths.

Although this is tangential to the major focus on the present investigation, the effects of the covariates on self-concept responses—after controlling for other variables—are of potential interest. The effects of age and SES were non-significant for all three domains. Although gender differences were small, girls scored significantly lower on all three self-concepts, but less so for Dutch self-concepts. Interestingly, students with parents born in a non-Western country scored slightly higher than those of students with parents from a Western country. This is consistent with the well-established “immigrant paradox”, in which immigrant students show better adaptation outcomes than their native-born counterparts (Fuligni, 1998, OECD, 2006, Sam, Berry, 2010) as demonstrated in previous Dutch research (van Geel & Vedder, 2010). However, the strongest effect of these background demographic variables is the negative effect of being in a higher academic track after controlling for individual student characteristics. This negative contextual effect is consistent with the well-established “big-fish-little-pond” effect (e.g., Marsh, Hau, 2003, Marsh et al, 2008, Nagengast, Marsh, 2012) in which students tend to have lower ASCs when they are in schools and classes in which the average achievement levels are higher than equally able students in schools and classes where the average achievement levels are lower. Hence, although further analyses of the effects of these covariates are beyond the scope of the present investigation, the effects are largely consistent with previous research. In summary, comparison of Models 4 and 4A (Table 5) demonstrates that support for DCT predictions is little affected by—or is even slightly stronger with—the inclusion of these covariates.

5. Discussion

Psychology researchers have long recognized that self-evaluations are formed in relation to social and temporal comparisons: the better I perform relative to others and relative to my past accomplishments, the more positive my self-evaluations. Based on the I/E model and its extension into DCT, we argue that internal dimensional comparisons among different domains are an additional frame of reference effect, with important implications for theory, policy-practice, and individual behavior. Indeed, a critical source of self-knowledge is knowing one's relative strengths and weaknesses in different domains, since this is the basis of many future academic choices, and influences competing strategies such as accurate self-evaluation, self-improvement, self-maintenance, and self-enhancement. The I/E model that underpins DCT is based mostly on how math and verbal achievements are related to MSC and VSC (Fig. 1A). Grounded on an impressive array of correlational, longitudinal, cross-cultural, experimental, and qualitative studies, there is clear support for theoretical predictions that better math achievement leads to lower VSCs and that better verbal achievement leads to lower MSCs—the critical, seemingly paradoxical prediction of the I/E model. In the present investigation, we explore new methodological, theoretical, and substantive insights into the extension of the I/E model into DCT.

Within a traditional construct validity perspective, it is typical to distinguish between convergent and discriminant validity. In validating ASCs in relation to achievement, convergent validity is represented by the horizontal paths in the I/E model (Fig. 1), whereas discriminant validity is related to the cross paths. In this respect, our results show that support for both the convergent and the discriminant validity of ASCs is stronger in relation to class marks than to standardized test scores. Using this same logic in reverse, it is also possible to argue that support for the convergent and discriminant validity of achievement measures is stronger for class marks than for test scores—at least in relation to ASCs. This is also evident in that correlations among the standardized test scores (.523 to .574) are much higher and less differentiated than correlations among the class marks (.157 to .335). Indeed, test scores are only modestly correlated with class marks (Mn $r = .272$) and no more highly correlated with class marks than ASCs (Mn $r = .296$). Thus, it is not surprising that the predictive validity of test scores dropped substantially when class marks were included in the prediction of ASCs (e.g., Model 4 in Table 3A, Table 3B, Table 4A, Table 4B). These results are consistent with our a priori predictions based on previous research, showing that class marks are more highly correlated with ASCs than test scores. Thus, class marks are a more “locally dominant” indicator of achievement, a more direct measure of achievement, and also more responsive to the motivational influences that it shares with ASC. Based on our findings, we argue that future DCT research—and educational research more generally—should more fully differentiate between class marks and standardized test scores as distinct indicators of achievement. In relation to ASC research in particular, class marks are a more useful measure of achievement than are standardized test scores. However, class marks also present greater challenges, in that they tend to be idiosyncratic in relation to specific subjects, individual teachers, and type of assessment (e.g., tests, homework, projects, classroom participation). Central to DCT is the expansion of the classic I/E model (Fig. 1A) to include more than just math and verbal domains (Fig. 1B). Previous I/E studies have mostly focused on math and verbal domains, maximally dissimilar school subjects in relation to the theoretical continuum of ASCs in Fig. 2. However, the theoretical basis of the I/E model and its extension to DCT posits that negative cross paths (contrast) associated with maximally dissimilar (far) domains will diminish and may even become positive (assimilation) for maximally similar (near) domains. In the present investigation we operationalized this extension and proposed new statistical tests of these predictions. Consistent with a priori predictions, far cross paths (math compared to English and Dutch) were consistently negative, whereas near cross paths (English compared to Dutch) were consistently much smaller (small positive, non-significant, or small negative). However, more research is needed to develop this extension more fully in relation to a wider array of academic subjects, and to other conditions that may prompt students to contrast or assimilate information from different domains in the formation of ASCs (also see related work on assimilation and contrast in social comparison research; e.g., Huguet et al, 2009, Suls, Wheeler, 2000). Furthermore, although there is a strong theoretical and empirical basis for this continuum—and support for it in the present investigation—it would be useful to extend this area of research further by asking students to directly rate the perceived similarity between different academic domains. This could be used to validate the continuum but also provide more nuanced tests of DCT predictions such that the perceived "nearness" or "farness" of different domains might influence the size of the observed relations over and above the more objective positioning of constructs on the continuum.

In social comparison research Huguet et al. (2009) introduced meVclass ratings to test social comparison processes. Extending this innovation, we posited that the use of meVclass ratings would allow us to disentangle social comparison and internal dimensional comparison processes, which are at the heart of DCT. In some respects these meVclass measures worked well, in that these ratings were surprisingly free of response biases; the grand means across all respondents were almost exactly equal to the average response category (see earlier discussion) and meVclass ratings had good convergent and discriminant validity in relation to ASCs, class marks, and test scores. However, empirical tests only partially supported our hypothesis that inclusion of meVclass ratings would substantially reduce or eliminate the horizontal paths that reflect external social comparisons but have little or no effect on cross paths that reflect internal dimensional comparisons. In particular, we found that the horizontal paths were still substantial (even though substantially reduced) and that negative cross paths became less negative (even though the reduction was much smaller than for the horizontal paths). Several features of the meVclass ratings and the supplemental analyses offer further insight into why support was only partial. Of critical importance, correlations among three (Dutch, English, math) meVclass ratings were surprisingly small (Mn $r = .030$), which leads us to suspect that internal dimensional comparison processes might influence meVclass ratings. Support for this ex post facto supposition comes from a series of models in which meVclass ratings (instead of ASCs) are regressed on measures of achievement. These showed that far cross paths from test scores, and particularly school marks, were significantly negative. These analyses with meVclass ratings as outcomes provide particularly strong support for the generalizability and robustness of a priori DCT predictions, generalizing to measures that would seem not to have even an implicit demand for internal comparisons. However, we also asked why the far cross paths should be negative for meVclass ratings—apparently reflecting an internal comparison process—when the actual items specifically asked them to report the result of the external comparison in a straightforward manner. In order to test the possibility that there is an internal dimensional comparison process in the school marks, we regressed class marks on the test scores.

Interestingly, we found that far cross paths relating test scores to class marks were also negative. Because different teachers assigned the class marks in each subject, the result is unlikely to reflect teacher strategies in assigning class marks. Our tentative suggestion is that the internal comparison process has motivational properties that are reflected in the class marks achieved in different school subjects. Thus, students might over-achieve in their relatively best subjects (through some combination of academic choice behavior, greater motivation, effort, and conscientiousness) and under-perform in their relatively worst subjects (through some combination of procrastination, academic choice behavior and reduced motivation, effort, and conscientiousness). We suspect that this explains in part why far cross paths are negative even when meVclass and class marks are the outcome variables (but not when test scores are the outcome—see Model 8). This explanation is also consistent with correlations among the class marks, and particularly among the meVclass ratings, being so much lower than those among the test scores. These speculations, if confirmed by subsequent research, will have important implications for DCT, frame of reference research, and the use of meVclass ratings as a (perhaps) not-so-pure measure of social comparison processes. However, the implications are even more important for understanding student motivation in the pursuit of mastering materials in what they perceive as their best and worst school subjects. Indeed, over time this sort of process could become a self-fulfilling prophecy. However, much as they are consistent with the results, these are post-hoc speculations requiring further work.

6. Limitations and directions for further research

An important contribution of the present investigation is that it provides a prototype for further research on the newly formulated DCT. Key features include the distinction between matching, far and near paths relating achievement and self-concepts in different domains, as well as new statistical tests that facilitate the evaluation of a priori predictions based on DCT. An obvious direction for further research is expanding the range of achievement domains to be considered. In particular, the inclusion of different science domains is important in terms of increasing emphasis in educational psychology on student development in science and selection of STEM courses. Fundamental to this expansion is the theoretical continuum of academic subjects along the math-to-verbal continuum (Fig. 2). Importantly, this continuum has a strong theoretical and empirical basis in the Marsh/Shavelson multidimensional, hierarchical model of self-concept. Nevertheless, further research is needed as to whether pair-wise differences between the perceived complementarity/antagonism of a whole range of academic domains can be adequately represented by this continuum in relation to DCT (e.g., Goetz et al., 2010 quantified differences between school subjects based on the academic emotions).

Support for the DCT, as in most I/E research more generally, is largely based on cross-sectional, correlational studies, so that causal interpretations can only be offered tentatively and interpreted cautiously. Fortunately, there is now a growing body of I/E studies, using various combinations of qualitative introspection studies and longitudinal, quasi-experimental, and true experimental designs, with random assignment of students, that support the I/E model (see earlier discussion). However, there is a need for further research along these lines to test the causal interpretation of new theoretical predictions based on DCT—particularly in relation to the distinction between matching, near, and far comparisons, but also in relation to conditions under which near comparisons are likely to result in assimilation rather than contrast.

Following Huguet et al. (2009), the meVclass ratings were designed to represent pure measures of social comparison based on single-item responses to each domain. Although the results were not entirely consistent with a priori predictions, we posed heuristic (ex post facto) speculations to explain these results, which have potentially profound implications for understanding student motivation in relation to what they perceive as their best and worst school subjects. Although this is clearly beyond the scope of the present investigation, it is a potentially important direction for further research based on multi-item measures of these constructs, which would facilitate stronger latent-variable models and a better understanding of the processes that they represent. Indeed, an important limitation of the current study is that the social comparison measures were all based on single items, which clearly suggests the need for further studies based on multi-item measures of these constructs to more precisely evaluate the extent to which our results may have been negatively impacted by these single items measures. Nevertheless, as a first direct empirical test of DCT, we note that the current results are highly promising. In this regard, we note that it would also be useful to have direct measures of the hypothesized internal and external comparisons, but results of our attempt to do this in the present investigation suggest that this might prove difficult.
A strength of the present investigation is that it was based on the Dutch secondary school system, which is so highly tracked. Specifically, this facilitated tests of a priori predictions about the differential effects of class marks (based substantially on how students compare with other students in their same class) and test scores (based on a common metric that is relatively independent of the class, teacher, or school). However, this strength is also a limitation, in that this separation is likely to be less clear in systems in which there is no tracking, or in which the tracking is not so clearly defined. Hence, there is a need to evaluate the generalizability of the results in other contexts.

7. Policy and practice implications

The contributions of the present investigation are primarily theoretical (testing new theoretical predictions based on DCT) and methodological (stronger designs and better statistical methodology for testing these hypotheses). However, the I/E model and its extension into DCT also has implications for teachers and classroom practice. When teachers, parents, and "significant others" are asked to infer students' self-concepts (see Dai, 2002, Marsh, 2007, Marsh, Craven, 1997), the responses apparently reflect the external comparison process mainly, such that inferences are not nearly as domain specific as self-responses by students themselves. The responses of teachers, parents, and others imply that, consistent with corresponding measures of achievement, students who are bright in one area tend to be seen as having good ASCs in all academic areas, whereas students with lower ability in one area are seen as having poor ASCs in all areas. However, if teachers and others better understood the formation of self-concepts in different academic domains, they would be able to understand their students better and provide more appropriate, credible feedback, particularly for less able students. Even bright students might have an average or below-average self-concept in their weakest school subjects; this is somewhat paradoxical, in relation to their good achievement (i.e., relative to other students but not relative to their own performance in other school subjects). In a similar vein, even poor students may have an average or above-average self-concept in their best school subject; this appears paradoxical on the basis of their below-average achievement in that subject (an external, social comparison), but not when considered in relation to their other school subjects (an internal, dimensional comparison). In the present investigation we add to these insights the distinction between near and far comparisons. Hence, feedback from teachers and others should particularly reinforce the complementarity of near domains, where accomplishments in one domain might have positive effects in other, similar domains. Teachers should also reinforce the complementarity of accomplishments in a student's best school subject with far school subjects, to undermine the negative contrast effects that undermine self-belief in relation to these subjects.

We also note that the distinction between class marks and standardized test scores has important implications for instructional practice, assessment practices, individual student self-concepts, and students' long-term choice behaviors. Particularly in the present investigation, where classes are highly tracked, these two indicators of achievement are only moderately correlated, so it is clear that class marks and test scores represent distinct constructs. Student self-concepts are substantially more highly correlated with class marks than with test scores, and many previous studies show that self-concepts are highly predictive of students' academic choices, aspirations, and long-term attainment. However, what is new in the present investigation is the finding that high class marks in one subject are even more detrimental to self-concepts in different domains than high test scores, particularly when based on far comparisons. There is also some suggestion that this becomes a self-fulfilling prophecy in that, relative to standardized achievement scores, students get better than expected class marks in their best subjects and, perhaps, lower than expected class marks in their weakest subjects. Although these interpretations are still highly speculative and require further research, this interpretation of our results is consistent with dimensional comparison processes, the main focus of the present investigation.

References

Albert, S. (1977). Temporal comparison theory. Psychological Review, 84, 485–503. doi:10.1037/0033-295X.84.6.485.

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. Psychological Science, 21, 174–177. doi:10.1177/0956797609357718.

Beauducel, A., & Herzberg, P. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Structural Equation Modeling, 13, 186–203.

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. Journal of Educational Psychology, 90, 102–110. doi:10.1037/0022-0663.90.1.102.

Brickman, P., & Bulman, R. J. (1977). Pleasure and pain in social comparison. In J. M. Suls & R. L. Miller (Eds.), Social comparison processes: Theoretical and empirical perspectives (pp. 149–186). Washington, DC: Hemisphere.

Centre on International Education Benchmarking. (no date). Netherlands. <http://www.ncee.org/programs-affiliates/center-on-international-education-benchmarking/top-performing-countries/netherlands-overview/netherlands-education-for-all/>. Accessed 11.05.14.

Chen, S., Yeh, Y., Hwang, F., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. Learning and Individual Differences, 23, 172–178.

Collins, R. (1996). For better or worse: The impact of upward social comparison on self-evaluations. Psychological Bulletin, 119, 51–69.

Dai, D. Y. (2002). Incorporating parent perceptions: A replication and extension study of the internal-external frame of reference model of self-concept development. Journal of Adolescent Research, 17, 617–645.

Dai, D. Y., & Rinn, A. N. (2008). The big-fish–little-pond-effect: What do we know and where do we go from here? Educational Psychology Review, 20, 283–317.

Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), Health, coping, and well-being: Perspectives from social comparison theory (pp. 329–358). Mahwah, NJ: Erlbaum.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. Structural Equation Modeling, 9, 327–346.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. British Journal of Mathematical and Statistical Psychology, 47, 309–326.

Eccles, J. S., Vida, M. N., & Barber, B. (2004). The relation of early adolescents' college plans and both academic ability and task-value beliefs to subsequent college enrollment. The Journal of Early Adolescence, 24(1), 63–77.

Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7, 117–140.

Fuligni, A. J. (1998). The adjustment of children from immigrant families. Current Directions in Psychological Science, 7, 99–103.

Goetz, T., Cronjaeger, H., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. Contemporary Educational Psychology, 35, 44–58. doi:10.1016/j.cedpsych.2009.10.001.

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic

emotions: Testing the internal/external frame of reference model for academic enjoyment. Contemporary Educational Psychology, 33, 9–33. doi:10.1016/ j.cedpsych.2006.12.002.
Goetz, T., Frenzel, A. C., Pekrun, R., & Hall, N. C. (2006). The domain specificity of academic emotional experiences. The Journal of Experimental Education, 75, 5–29. doi:10.3200/JEXE.75.1.5-29.
Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. Self and Identity, 1–16.
Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. London: SAGE Publications.
Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/ performance measures. Review of Educational Research, 52, 123–142.
Huguet, P., Dumas, F., Marsh, H. W., Regner, I., Wheeler, L., Suls, J., et al. (2009). Clarifying the role of social comparison in the Big-Fish-Little-Pond Effect (BFLPE): An integrative study. Journal of Personality and Social Psychology, 97, 671–710.
Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. In J. R. Nesselroade & B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 303–351). New York: Academic Press.
Kuyper, H., Dijkstra, P., Buunk, A. P., & Van der Werf, M. P. C. (2011). Social comparisons in the classroom: An investigation of the better than average effect among secondary school children. Journal of School Psychology, 49, 25–53.
Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. American Educational Research Journal, 23, 129–149.
Marsh, H. W. (1987). The big-fish–little-pond effect on academic self-concept. Journal of Educational Psychology, 79, 280–295.
Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson Model. Journal of Educational Psychology, 82, 623–636.
Marsh, H. W. (1991). Failure of high ability schools to deliver academic benefits commensurate with their students' ability levels. American Educational Research Journal, 28, 445–480.
Marsh, H. W. (2007). Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology. Leicester, UK: British Psychological

Society.
Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., et al. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. Journal of Educational Psychology, 105(1), 108–128.
Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. Journal of Educational Psychology, 80, 366–380. doi:10.1037/0022-0663.80.3.366.
Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), Handbook of classroom assessment: Learning, achievement, and adjustment (pp. 131–198). Orlando, FL: Academic Press.
Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. Perspectives on Psychological Science, 1, 133–163. doi:10.1111/j.1745-6916.2006.00010.x.
Marsh, H. W., & Hau, K. T. (2003). Big-Fish – Little-Pond effect on academic selfconcept: A cross-cultural (26-country) test of the negative effects of academically selective schools. American Psychologist, 58(5), 364.
Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internalexternal frame of reference predictions across 26 countries. Journal of Educational Psychology, 96, 56–67. doi:10.1037/0022-0663.96.1.56.
Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? Journal of Experimental Education, 64, 364–390.
Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of Educational Psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. International Journal of Testing, 6, 311–360. doi:10.1207/s15327574ijt0604_1.
Marsh, H. W., Kong, C. K., & Hau, K. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. Journal of Educational Psychology, 93, 543–553. doi:10.1037/0022-0663.93.3.543.
Marsh, H. W., Möller, J., Parker, P., Xu, M. K., Benjamin, N., & Pekrun, R. (2014).

Internal/external frame of reference model. In J. D. Wright (Ed.), The international encyclopedia of the social and behavioral sciences (2nd ed.). Oxford, UK: Elsevier. In press.

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Regner, I., et al. (2010). Phantom behavioral assimilation effects: Systematic biases in social comparison choice studies. Journal of Personality, 78, 671–710. doi:10.1111/j.1467-6494.2010.00630.x.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., et al. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20, 319–350. doi:10.1007/s10648-008-9075-6.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20, 107–125.

Marsh, H. W., & Yeung, A. S. (2001). An extension of the internal/external frame of reference model: A response to Bong (1998). Multivariate Behavioral Research, 36(3), 389–420.

Meelissen, M., & Punter, A. (2012). The Netherlands. In I. V. S. Mullis, M. O. Martin, C. A. Minnich, G. M. Stanco, A. Arora, V. A. S. Centurino, et al. (Eds.), TIMSS 2011 encyclopedia: Education policy and curriculum in mathematics and science (Vol. 2, pp. 619–630). TIMSS & PIRLS International Study Center, Boston College.

Möller, J. (2005). "Paradoxical" effects of praise and criticism: Social, dimensional, and temporal comparisons. British Journal of Educational Psychology, 75, 275–295. doi:10.1348/000709904X24744.

Möller, J., & Husemann, N. (2006). Internal comparisons in everyday life. Journal of Educational Psychology, 98, 342–353. doi:10.1037/0022-0663.98.2.342.

Möller, J., & Köller, O. (2001a). Dimensional comparisons: An experimental approach to the Internal/External frame of reference model. Journal of Educational Psychology, 93, 826–835. doi:10.1037/0022-0663.93.4.826.

Möller, J., & Köller, O. (2001b). Frame of reference effects following the announcement of exam results. Contemporary Educational Psychology, 26, 277–287. doi:10.1006/ceps.2000.1055.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. Psychological Review, 120(3), 544–560. doi:10.1037/a0032459.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the Internal/External frame of reference model of academic achievement and academic self-concept. Review of Educational Research, 79, 1129–1167. doi:10.3102/0034654309337522.

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal I/E model: An integration of models of relations between academic achievement and self-concept. American Educational Research Journal, 48, 1315–1346. doi:10.3102/0002831211419649.

Möller, J., & Savyon, K. (2003). Not very smart, thus moral: Dimensional comparisons between academic self-concept and honesty. Social Psychology of Education, 6, 95–106.

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. European Journal of Psychology of Education, 21, 467–487. doi:10.1007/BF03173515.

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical & Statistical Psychology, 38, 171–189.

Muthén, L. K., & Muthén, B. O. (2008–2013). Mplus user's guide. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. Journal of Educational Psychology, 104(4), 1033–1053. doi:10.1037/a0027697.

Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences (pp. 115–143). doi:10.1037/11706-004.

Oehlert, G. W. (1992). A note on the delta method. American Statistician, 46(1), 27–29.

OECD (2006). Where immigrant students succeed—A comparative review of performance and engagement in PISA 2003. Paris: OECD.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2013).

Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. Educational Psychology, 34, 29–48. doi:10.1080/ 01443410.2013.797339.
Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. Developmental Psychology, 48(6), 1629–1642. doi:10.1037/a0029167.
Pekrun, R. (2006). The control–value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educational Psychology Review, 18, 315–341. doi:10.1007/s10648-006-9029-9.
Pinxten, M., De Fraine, B., Van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. British Journal of Educational Psychology, 80, 689–709.
Pohlmann, B., & Möller, J. (2009). On the benefits of dimensional comparisons. Journal of Educational Psychology, 101, 248–258. doi:10.1037/a0013151.
Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. Structural Equation Modeling, 11, 621–637. doi:10.1207/s15328007sem1104_7.
Sam, D. L., & Berry, J. W. (2010). Acculturation when individuals and groups of different cultural backgrounds meet. Perspectives on Psychological Science, 5(4), 472–481. doi:10.1177/1745691610373075.
Scheerens, J., Luyten, H., & van Ravens, J. (2011). Perspectives on educational quality: Illustrative outcomes on primary and secondary schooling in the Netherlands. New York: Springer.
Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., et al. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. British Journal of Social Psychology, 47(1), 73–103. doi:10.1348/014466607X202309.
Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. Review of Educational Research, 46, 407–444.
Suls, J. E., & Wheeler, L. E. (2000). Handbook of social comparison: Theory and research. Kluwer Academic Publishers.
Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), Advances in experimental social psychology: Social psychological

studies of the self: Perspectives and programs (Vol. 21, pp. 695–709).

van Geel, M., & Vedder, P. (2010). The adaptation of non-western and Muslim immigrant adolescents in the Netherlands: An immigrant paradox? Scandinavian Journal of Psychology, 51, 398–402.

Xu, M. K., Marsh, H. W., Hau, K.-T., Ho, I. T., Morin, A. J. S., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. Journal of Educational Psychology, 105(2), 489–503. doi:10.1037/a0031333.

Zell, E., & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. Journal of Personality and Social Psychology, 97, 467–482.

Table 1
Latent correlations among constructs considered in subsequent models.

	Self-concepts			Class marks			Test scores			meVclass ratings			
	1	2	3	4	5	6	7	8	9	10	11	12	13
Self-concept factors													
(1) Dutch	1.0												
(2) English	.190	1.0											
(3) Math	−.060	−.169	1.0										
Class marks													
(4) Dutch	.602	.102	.068	1.0									
(5) English	.141	.717	−.054	.335	1.0								
(6) Math	−.018	−.114	.759	.285	.157	1.0							
Test scores													
(7) Dutch	.234	.210	.013	.253	.240	.106	1.0						
(8) English	.061	.327	.062	.114	.282	.089	.519	1.0					
(9) Math	−.025	.072	.328	.080	.109	.282	.523	.574	1.0				
meVclass comparison ratings													
(10) Dutch	.705	.176	.024	.479	.171	.072	.198	.078	.031	1.0			
(11) English	.078	.826	−.123	.039	.602	−.082	.133	.278	.059	.145	1.0		
(12) Math	−.070	−.177	.854	.062	−.054	.680	−.011	.037	.287	.046	−.101	1.0	
(13) Global	.234	.210	.197	.221	.216	.197	.136	.112	.133	.294	.214	.230	1.0

Note: Correlations are based on Model CFA3, described in more detail in the Supplemental Materials. Self-concepts in the three subjects are latent variable factors based on multiple indicators (see factor loadings in Supplemental Materials) while all other variables are manifest variables based on single indicators. Comparison ratings are based on student ratings of how they compare with other students in their class in each subject, and globally (meVclass).

Table 2
Average correlations (based on values from Table 1) designed to test a priori predictions: mean of standardized path coefficients and standard errors.

	Estimate	S.E.	t-Value
Mean correlations within methods[a]			
Self-concept (self)	−0.013	0.008	−1.643
Class marks (mark)	0.259	0.010	27.063**
Test scores (test)	0.539	0.014	39.828**
meVclass (MvC)	0.030	0.007	4.569**
Pair-wise differences mean correlations within methods			
Self-mark	−0.272	0.010	−28.221**
Self-test	−0.552	0.015	−37.452**
Self-meVclass	−0.043	0.007	−5.948**
Mark-test	−0.280	0.016	−17.380**
Mark-meVclass	0.229	0.010	23.251**
Test-meVclass	0.508	0.014	36.307**
Mean convergent validities[b]			
Self and mark	0.693	0.006	113.002**
Self and test	0.296	0.008	36.511**
Self and meVclass	0.795	0.004	210.770**
Mark and test	0.272	0.010	26.009**
Mark and meVclass	0.587	0.006	92.981**
Test and meVclass	0.254	0.007	38.870**
Pairwise differences in mean convergent validities			
Self-mark vs. self-test	0.397	0.009	43.432**
Self-mark vs. self-meVclass	−0.103	0.006	−16.055**
Self-mark vs. mark-test	0.420	0.011	39.352**
Self-mark vs. mark-meVclass	0.106	0.004	23.762**
Self-mark vs. test-meVclass	0.438	0.008	52.388**
Self-test vs. self-meVclass	−0.499	0.009	−56.780**
Self-test vs. mark-test	0.024	0.009	2.662**
Self-test vs. mark-meVclass	−0.291	0.010	−29.790**
Self-test vs. test-meVclass	0.042	0.006	7.630**
Self-meVclass vs. mark-test	0.523	0.011	49.496**
Self-meVclass vs. mark-meVclass	0.208	0.006	35.133**
Self-meVclass vs. test-meVclass	0.541	0.007	76.975**
Mark-test vs. mark-meVclass	−0.315	0.011	−28.778**
Mark-test vs. test-meVclass	0.018	0.010	1.790
Mark-meVclass vs. test-meVclass	0.333	0.009	39.045**

Note: Using the model constraint procedure in Mplus, the mean correlation across the three domains was computed as part of the model used to estimate correlations in Table 1 (for more detail, see discussion of Model CFA2: Supplemental Materials). Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

[a] Mean correlations within methods refer to the average correlation across the three domains (Dutch, English, math) for each type of measure (self-concept, class marks, test scores and meVclass ratings). Differences refer to pair-wise comparisons of the means.

[b] Mean convergent validities refer to the average correlation between scores based on matching domains but different methods (e.g., agreement among self-concept ratings and class marks when the domains are matching). Differences refer to pair-wise comparisons of the means.

** $p < .01$.

Table 3A

Five models predicting three academic self-concept factors (Dutch, English, and Math) based on alternative sets of predictor variables: standardized path coefficients as a ratio of standard errors relating achievement measures to ASC factors (Dutch, English, Math).

Predictor variables	Model 1	Model 2	Model 3	Model 4	Model 5
	Predictions of Dutch self-concept				
Dutch: Test	.340/.016			.161/.014	.104/.011
English: Test	.001/.018			−.001/.015	−.003/.013
Math: Test	−.203/.017			−.102/.014	−.078/.012
Dutch: Mark		.678/.015		.645/.015	.401/.014
English: Mark		−.055/.011		−.075/.012	−.092/.013
Math: Mark		−.202/.010		−.178/.011	−.142/.014
Dutch: meVclass			.699/.012		.508/.012
English: meVclass			−.050/.008		.014/.011
Math: meVclass			−.123/.008		−.011/.012
Global meVclass			.068/.010		.040/.009
Mult R Sq	.085/.007	.404/.013	.513/.010	.421/.012	.627/.009
	Predictions of English self-concept				
Dutch: Test	.121/.014			.021/.011	.032/.008
English: Test	.388/.019			.168/.014	.074/.010
Math: Test	−.214/.015			−.051/.012	−.024/.009
Dutch: Mark		−.116/.010		−.108/.010	−.059/.009
English: Mark		.784/.011		.738/.011	.378/.012
Math: Mark		−.208/.009		−.202/.009	−.102/.010
Dutch: meVclass			.051/.006		.050/.007
English: meVclass			.797/.009		.553/.012
Math: meVclass			−.110/.007		−.031/.009
Global meVclass			.050/.007		.026/.006
Mult R Sq	.137/.010	.575/.011	.697/.008	.597/.010	.782/.006
	Predictions of Math self-concept				
Dutch: Test	−.179/.013			−.109/.009	−.040/.008
English: Test	−.131/.016			−.018/.011	−.002/.009
Math: Test	.497/.015			.202/.011	.091/.008
Dutch: Mark		−.116/.009		−.091/.009	−.054/.008
English: Mark		−.142/.008		−.133/.009	−.070/.009
Math: Mark		.815/.009		.762/.010	.374/.012
Dutch: meVclass			−.014/.006		.011/.007
English: meVclass			−.039/.006		.007/.008
Math: meVclass			.848/.008		.571/.011
Global meVclass			.014/.006		.008/.006
Mult R Sq	.153/.008	.614/.010	.731/.007	.643/.009	.802/.006

Note: In Models 1–5, three latent academic self-concept factors (Dutch, English, Math) were predicted from corresponding sets of test scores (test), class marks (mark), or meVclass ratings. Thus, for example, in Model 1 all three self-concept factors were predicted by Dutch, English and Math Test scores, resulting in 9 path coefficients. Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

Table 3B

Average path coefficients (based on values from Table 3A) designed to test a priori predictions: mean of standardized path coefficients as a ratio of standard errors.

	Model 1	Model 2	Model 3	Model 4	Model 5
Mean horizontal and cross paths (averaged across domains)					
Paths from test scores					
Horizontal paths	.408/.011			.177/.007	.090/.006
Cross paths	−.101/.006			−.043/.004	−.019/.003
Diff horizontal-cross	.509/.015			.220/.010	.109/.007
Near paths	.061/.011			.010/.009	.015/.007
Far paths	−.182/.008			−.070/.006	−.036/.005
Diff near-far: test	.243/.015			.080/.012	.051/.010
Paths from class marks					
Horizontal paths		.759/.008		.715/.008	.385/.008
Cross paths		−.140/.004		−.131/.005	−.087/.004
Diff horizontal-cross		.896/.009		.846/.009	.471/.011
Near paths		−.086/.008		−.092/.008	−.076/.008
Far paths		−.167/.005		−.151/.005	−.092/.006
Diff near-far: test		.081/.009		.059/.010	.016/.011
Paths from meVclass ratings					
Horizontal paths			.781/.006		.544/.007
Cross paths			−.048/.003		.007/.004
Diff horizontal-cross			.829/.006		.537/.008
Near paths			.000/.006		.032/.006
Far paths			−.072/.004		−.006/.005
Diff near-far: test			.072/.006		.038/.008
Global meVclass			.044/.005		.025/.005
Differences in paths based on test score and class marks					
Horizontal paths				−.538/.012	−.295/.010
Cross paths				.088/.007	.067/.006
Diff horizontal-cross				−.626/.016	−.362/.013
Near paths				.102/.014	.090/.012
Far paths				.081/.009	.056/.008
Diff near-far: test				.021/.018	.034/.016

Note: Using the model constraint procedure in Mplus, the mean path coefficient across the three domains was computed for paths from Table 3A (also see Figure 1): Horizontal paths between matching ASCs and predictor variables; Cross paths between non-matching ASCs and predictor variables; Near cross paths (relating English and Dutch domains to each other); Far cross paths (paths relating English and Dutch domains to the Math domain). Also shown are tests of the difference between paths based on test scores and class marks and between near and far cross paths. Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

Table 4A
Five models predicting meVclass, class marks and test scores in three domains (Dutch, English, and Math) based on alternative sets of predictor variables: standardized path coefficients as a ratio of standard errors.

Predictors	Outcome measures				
	meVclass Model 6A	meVclass Model 6B	meVclass Model 6C	Class marks Model 7	Test scores Model 8
	Predictions of Dutch outcomes				
Dutch: Test	.247/.013		.109/.013	.289/.015	
English: Test	.009/.015		−.001/.015	.007/.019	
Math: Test	−.103/.014		−.045/.013	−.075/.016	
Dutch: Mark		.494/.013	.472/.012		.188/.017
English: Mark		.017/.010	.002/.011		.173/.015
Math: Mark		−.072/.010	−.062/.010		.025/.013
Mult R Sq	.046/.007	.234/.010	.242/.010	.068/.007	.092/.009
	Predictions of English outcomes				
Dutch: Test	.035/.012		−.037/.010	.170/.015	
English: Test	.354/.016		.167/.013	.272/.017	
Math: Test	−.103/.014		−.036/.012	−.136/.017	
Dutch: Mark		−.146/.010	−.141/.010		.010/.019
English: Mark		.674/.010	.638/.009		.272/.016
Math: Mark		−.146/.009	−.142/.009		.044/.016
Mult R Sq	.093/.009	.412/.009	.430/.009	.103/.009	.082/.009
	Predictions of Math outcomes				
Dutch: Test	−.182/.012		−.122/.010	−.028/.014	
English: Test	−.130/.015		−.030/.011	−.099/.018	
Math: Test	.457/.013		.195/.011	.354/.015	
Dutch: Mark		−.100/.010	−.073/.010		−.024/.017
English: Mark		−.135/.008	−.120/.008		.074/.016
Math: Mark		.730/.009	.680/.008		.277/.015
Mult R Sq	.128/.008	.497/.009	.520/.008	.088/.008	.084/.008

Note: In each of the first three models (6A–6C), meVclass ratings in three domains (Dutch, English, Math) are predicted from the set of corresponding three test scores (6A), three class marks (6B), or the combined sets of test scores and class marks (6C). In Model 7, the three class marks (Dutch, English, Math) were predicted from corresponding sets of test scores. In Model 8, the three test scores (Dutch, English, Math) were predicted from corresponding sets of class marks. Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

Table 4B
Average path coefficients (based on values from Table 4A) designed to test a priori predictions: mean of standardized path coefficients as a ratio of standard errors.

	Model 6A	Model 6B	Model 6A	Model 7	Model 8
Mean horizontal and cross paths (averaged across domains)					
Paths from test scores					
Horizontal paths	.352/.009		.157/.007	.305/.010	.246/.008
Cross paths	−.089/.005		−.045/.004	−.027/.007	.050/.007
Diff horizontal-cross	.442/.013		.202/.010	.331/.013	.195/.007
Near paths	.022/.010		−.019/.009	.089/.013	.092/.011
Far paths	−.144/.007		−.058/.006	−.084/.008	.030/.008
Diff near-far	.166/.013		.039/.012	.173/.015	.062/.011
Paths from class marks					
Horizontal paths		.632/.007	.597/.007		
Cross paths		−.097/.004	−.089/.005		
Diff horizontal-cross		.729/.009	.686/.009		
Near paths		−.064/.008	−.070/.008		
Far paths		−.113/.005	−.099/.005		
Diff near-far		.049/.009	.029/.009		
Difference in paths based on test score and class marks					
Horizontal paths			−.440/.012		
Cross paths			.044/.007		
Diff horizontal-cross			−.484/.015		
Near paths			.051/.014		
Far paths			.041/.009		
Diff near-far			.010/.017		

Note: Using the model constraint procedure in Mplus, the mean path coefficient across the three domains was computed for paths in Table 4A (also see Figure 1): Horizontal paths between matching ASCs and predictor variables; Cross paths between non-matching ASCs and predictor variables; Near cross paths (relating English and Dutch domains to each other); Far cross paths (paths relating English and Dutch domains to the Math domain). Also shown are tests of the difference between paths based on test scores and class marks and between near and far cross paths. Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

Table 5
Five models predicting three academic self-concept factors (Dutch, English, and Math) based on alternative sets of predictor variables: standardized path coefficients as a ratio of standard errors relating achievement measures to ASC factors (Dutch, English, Math).

Predictor variables	Model 4	Model 4A
	Predictions of Dutch self-concept	
Dutch: Test	.161/.014	.207/.015
English: Test	−.001/.015	.032/.016[a]
Math: Test	−.102/.014	−.062/.016
Dutch: Mark	.645/.015	.648/.015
English: Mark	−.075/.012	−.085/.012
Math: Mark	−.178/.011	−.180/.011
Track		−.121/.021
Female		−.032/.009
Age		.012/.009
SES		−.018/.011
Ethnic		−.075/.012
Mult R Sq	.421/.012	.431/.012
	Predictions of English self-concept	
Dutch: Test	.021/.011	.090/.011[a]
English: Test	.168/.014	.208/.014
Math: Test	−.051/.012	−.051/.013
Dutch: Mark	−.108/.010	−.087/.010
English: Mark	.738/.011	.720/.011
Math: Mark	−.202/.009	−.208/.009
Track		−.109/.016
Female		−.123/.008
Age		−.004/.007
SES		.003/.008
Ethnic		−.065/.009
Mult R Sq	.597/.010	.618/.010
	Predictions of Math self-concept	
Dutch: Test	−.109/.009	−.058/.010
English: Test	−.018/.011	.010/.012
Math: Test	.202/.011	.224/.013
Dutch: Mark	−.091/.009	−.076/.009
English: Mark	−.133/.009	−.149/.009
Math: Mark	.762/.010	.757/.010
Track		−.100/.015
Female		−.087/.007
Age		−.004/.007
SES		.004/.008
Ethnic		−.073/.009
Mult R Sq	.643/.009	.657/.000

Note: In Models 4 (repeated from Table 3A) and 4A, three latent academic self-concept factors (Dutch, English, Math) were predicted from corresponding sets of test scores (test) and class marks (mark). The models differ in that background covariates were added in Model 4A: SES (based on parent's education); Track (academic track; higher values represent more advance tracks); Ethnic (0 = nonWestern, 1 = Western). Coefficients that are more than twice the size of their SE are statistically significant ($p < .05$).

[a] These near path coefficients were nonsignificant in Model 4, but were statistically significant in Model 4.