JOURNAL OF
RESEARCH on ADOLESCENCE

JOURNAL OF RESEARCH ON ADOLESCENCE, 20(2), 482–506

DOI: 10.1111/j.1532-7795.2010.00644.x

The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories?

Gabriel Nagy
Max Planck Institute for Human Development

Helen M. G. Watt
Monash University

Jacquelynne S. Eccles
University of Michigan

Ulrich Trautwein, Oliver Lüdtke, and Jürgen Baumert
Max Planck Institute for Human Development

Gender differences in the development of children's and adolescents' academic self-perceptions have received increasing attention in recent years. This study extends previous research by examining the development of mathematics self-concept across grades 7–12 in three cultural settings: Australia (Sydney; $N = 1{,}333$), the United States (Michigan; $N = 2{,}443$), and Germany (four federal states; $N = 4{,}688$). Results of latent growth curve models document very similar patterns of self-concept development in males and females in the three settings. First, gender differences in favor of boys were observed at the beginning of the observation period (grade 7). Second, gender was not significantly related to self-concept change in either group, meaning that initial differences persisted across time. Third, the results provided no evidence that the form of the longitudinal change trajectories for mathematics self-concept differed across the cultural settings. This pattern of results is inconsistent with explanatory models that predict converging or diverging gender differences in mathematics self-concept. Furthermore, the results indicate that self-concept development may be highly similar across western cultural settings.

Requests for reprints should be sent to Gabriel Nagy, Center for Educational Research, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: nagy@mpib-berlin.mpg.de

Students' self-perceptions of their competence or ability are at the core of several psychological theories aimed at explaining learning gains and achievement-related choices. The development of children's and adolescents' academic self-views is thus of great interest to researchers working in the framework of these theories. Various labels have been used to describe competence self-perceptions (e.g., Pedhazur & Schmelkin, 1991); in this article, we use the term ''academic self-concept'' (or simply ''self-concept''). The domain-specific academic self-concept reflects an individual's evaluation of his or her ability in a specific domain or academic area. It is usually assessed via self-report measures. Typical self-concept items are ''I am quite good at mathematics'' (mathematics self-concept) and ''I have a poor vocabulary'' (verbal self-concept). Domain-specific academic self-concepts are characterized by rather descriptive self-evaluations (compared with the more affective self-evaluations used in global self-esteem or interest measures).

According to Eccles' expectancy-value theory (Eccles et al., 1983), academic self-concept is crucial to motivation and is a key determinant of task choice. Domain-specific self-concept—along with other constructs, such as task values—has been found to be highly related to achievement, even after controlling for prior achievement (e.g., Eccles, Wigfield, Harold, & Blumfeld, 1993; Trautwein, Lüdtke, Köller, & Baumert, 2006), and to predict current and future activity choices (e.g., Eccles & Wigfield, 1995; Feather, 1988; Trautwein, Lüdtke, Schnyder, & Niggli, 2006). Research conducted in the framework of Marsh's self-concept theory has produced similar results. Academic self-concept has been shown to be a stronger predictor of subsequent achievement than vice versa (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Marsh & Yeung, 1997a), and self-concept repeatedly has been found to predict course choices in high school (Marsh & Yeung, 1997b; Nagy et al., 2008) and future career aspirations (Nagy, Trautwein, Köller, Baumert, & Garrett, 2006) over and above achievement.

At the same time, numerous studies have revealed consistent, domain-specific gender differences in self-concept. Males are typically found to have higher self-concepts in mathematics and sports, in particular; females tend to have higher self-competence perceptions in languages and the arts (Marsh & Yeung, 1998). Such gender differences in academic self-concept at least partially explain gender differences in domain-specific school- and career-related choices (Nagy et al., 2006, 2008).

Given the close connection of academic self-concept to achievement and achievement-related choices, and given the potential of this construct to explain gendered patterns of choice behavior, the widespread interest in gender differences in self-concept development is hardly surprising. The literature offers two competing theories of self-concept development in males and females. On the one hand, *gender intensification* theory (Hill & Lynch, 1983) suggests that gender-role activities become more important to young adolescents over time as they try to conform more to gender-role stereotypes (see also Eccles, 1987). Females become more negative about male-stereotyped domains, such as mathematics, and males become more negative about female-stereotyped domains, such as arts. On the other hand, the *gender convergence* hypothesis proposes the opposite development in gender differences in self-concept (Fredricks & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Proponents of this perspective argue that boys begin school with higher and unrealistic expectations, but that they adjust these expectations over the school years in response to performance feedback and the increased salience of social comparison processes. According to this view, the gender gap in self-concept decreases because girls do not initially overestimate their capabilities to the same extent as boys, meaning that the (negative) change in girls is smaller or even absent.

The gender intensification and gender convergence hypotheses are formulated on a general level and do not consider contextual factors that might interact with self-concept development. However, psychological theory and empirical evidence point to the importance of contextual factors for understanding individuals' development. According to Eccles's expectancy-value theory, self-concept development depends on the cultural context to which individuals are exposed. The cultural context provides norms, such as gender-role stereotypes, which impact individuals' self-concept formation. These influences are thought to be transmitted by socializers' beliefs and behaviors (e.g., Fredricks & Eccles, 2002; Frome & Eccles, 1998) and by their direct effects on individuals' perceptions.

Whereas expectancy-value theory is primary concerned with the role of culturally anchored norms, other contextual factors, such as the school systems implemented, might also play a role. For example, first findings suggest that students' domain-specific self-concepts are more strongly determined by ipsative-like comparison processes in the German school system than in the school system implemented in the United States (Nagy et al., 2008).

Examinations of gender effects on self-concept development have rarely investigated cultural differences. It is thus not clear to what extent the empirical results presented to date (see below) generalize across cultures that differ in their educational systems and prevailing gender norms. This article makes a first attempt to overcome previous limitations by drawing on large and representative data sets to investigate the development of students' self-concept in relation to gender in three distinct cultural settings: Australia, Germany, and the United States. We focus on self-concept in the domain of mathematics and on the period from grades 7 to 12.

PREVIOUS FINDINGS ON STUDENTS' SELF-CONCEPT DEVELOPMENT

Early studies conducted in the 1980s and 1990s found that academic self-concept generally declines across middle childhood and early adolescence (e.g., Alexander & Entwisle, 1988; Eccles et al., 1993; Marsh, 1989; Wigfield et al., 1997). However, these studies also showed that the pattern of decline varies across academic domains, with the sharpest drop in mathematics and a more modest decline in English (e.g., Eccles et al., 1983; Eccles & Midgley, 1989; Marsh, 1989; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). Early research linking the age-graded pattern of self-concept to gender revealed increasing gender differences, suggesting that these differences emerge during early adolescence and grow larger during late adolescence (Eccles, 1987; Eccles, Adler, & Meece, 1984). However, later research has indicated that gender differences in self-concept emerge during the elementary school years and remain stable later on (Eccles et al., 1993; Marsh, 1989, 1993; Marsh & Yeung, 1998; Wigfield et al., 1997).

Recent studies were able to capitalize on more extended longitudinal assessments and more sophisticated statistical models, such as growth-curve modeling. For example, Fredricks and Eccles (2002) and Jacobs et al. (2002) drew on a data set spanning grades 1–12. They found a decreasing trend for mathematics, English, and sports self-concept. Initial gender differences in mathematics in favor of boys declined over time; gender differences in English in favor of girls increased until grade 8 and then declined somewhat; gender differences in sports in favor of boys declined very slightly. Most recently, Watt (2004) investigated the development of students' talent perceptions—a construct that is related to, but nevertheless different from, self-concept (Watt, 2000)—in mathematics and English. Watt found that gender differences in mathematics in favor of boys remained stable from grades 7 to 11, whereas there were no gender differences for English.

Taken together, the studies by Fredricks and Eccles (2002), Jacobs et al. (2002), and Watt (2004) paint a somewhat mixed picture of the development of self-concept in relation to gender. Whereas Fredricks and Eccles (2002) and Jacobs et al. (2002) found a pattern of gender convergence that was most evident for the domain of mathematics, Watt (2004) found such a pattern only for English (although point estimates for gender groups at each grade were not statistically significantly different) and observed a stable gender gap for mathematics. Because these studies are based on data sets from different cultural contexts with different school structures (i.e., the United States and Australia), it might be inferred that the cultural/institutional setting moderates students' self-concept development in relation to gender. Such an interpretation is very appealing, but the difference between studies may equally be attributable to other artifacts.

First, the studies used measures based on different theoretical frameworks. Whereas the data set gathered in the United States (i.e., Fredricks & Eccles, 2002; Jacobs et al., 2002) assessed students' competence beliefs, Watt (2004) used a measure of perceived talent. As noted by Watt (2000), self-concept and perceived competence are related, but load on different factors in factor analysis. Second, the growth curve modeling approaches (Meredith

& Tisak, 1990) used differed in the degree of model restriction. Fredricks and Eccles (2002) and Jacobs et al. (2002) used a growth model with linear and quadratic time trends; Watt's model (2004) included an additional cubic trend. Third, the studies covered different time frames. Whereas the longitudinal studies conducted in the United States used observations from grades 1 to 12, the Australian study relied on assessments from grades 7 to 11. It is thus possible these methodological and analytical differences masked common aspects in self-concept development across the studies.

Cross-cultural comparisons overcoming these limitations are required to draw firmer conclusions on gender differences in self-concept development. Furthermore, the scope of investigations of cultural/institutional differences in gender-specific development of self-concept should be extended to cover further settings. Although the United States and Australia differ from one another in many respects, these settings also share many features, such as a common language and common aspects of the school curriculum (see below). In order to gain a deeper understanding of how contextual influences impact gender differences in self-concept development, we therefore extended previous research by including a new setting with a rather different educational structure from Australia and the United States, namely, Germany.

DIFFERENCES IN GENDER NORMS AND EDUCATIONAL SYSTEMS ACROSS CULTURAL SETTINGS CONSIDERED

Individual development cannot be understood in isolation from the personal and cultural context. Two aspects of cultural differences seem especially important when examining students' self-concept development in relation to gender. First, culture-specific gender norms operating on the macrosocietal level (Williams & Best, 1990) might promote or inhibit gender differentiated developmental trajectories. Second, differences in the structure of educational systems might foster or inhibit gender effects on self-concept development (e.g., Nagy et al., 2008).

Cultures vary in the degree to which sex roles are emphasized (e.g., Costa, Terracciano, & McCrae, 2001). The degree of gender differences in sex-typed attributes, such as mathematics self-concept, is likely to depend on the extent to which gender norms are differentiated in a cultural setting (Lynn & Martin, 1997). Gender norms are said to be highly differentiated when many adjectives are clearly ascribed to men or to women, but not to both. Gender differentiation is said to be low when adjectives are ascribed to both genders (Williams & Best, 1990). As shown by Williams and Best, countries differ in their degree of gender norm differentiation. Regarding the cultural settings investigated in this study (i.e., Australia, Germany, and the United States) Williams and Best (1990, p. 251) reported relatively high gender norm differentiation in Germany and substantially lower gender norm differentiation in Australia and the United States.

The specific affordances and constraints of national educational systems are of great relevance to differential developmental processes and their explanation (Schnabel, Alfeld, Eccles, Köller, & Baumert, 2002). The school systems under investigation differ in a number of respects. First, tracking takes place on the within-school level in Australia and the United States, but on the between-school level in Germany (with students being assigned to the lower, intermediate, or academic track). Second, tracking is implemented on the level of academic domains in Australia and the United States, but not in Germany, where students are assigned to tracks depending on their "overall" achievement in primary school. Third, tracking starts at different ages, beginning earlier in Germany (most often in grade 5) than in Australia or the United States. Furthermore, the German school tracks themselves differ. The intensity of the curriculum differs markedly across three secondary school tracks, meaning that once tracking has taken place students are exposed to different curricular demands for the rest of their school career. In Australia and the United States, in contrast, exposure to curricular demands is more flexible in terms of reversibility and domain specificity. It can be thus concluded that the German school system is rather different from the systems implemented in the United States and Australia (for more details, see Appendix A).

In sum, the cultural settings considered in the present investigation differ in terms of the prevailing gender norms (Williams & Best, 1990) as well as in the structure of the educational systems implemented (Schnabel et al., 2002). Whereas Australia and the United States appear to be relatively similar to each other in both respects, Germany is characterized by more differentiated gender norms and by a more rigid school structure.

THE PRESENT INVESTIGATION

The present investigation was undertaken to provide more conclusive evidence on how cultural and institutional influences impact adolescents' self-concept development. To this end, we focused on one domain-specific academic self-concept, namely, mathematics self-concept. There were several reasons for this choice. First, mathematics is probably one of the best investigated domains of self-concept and remains the focus of many researchers. Second, mathematics is a highly sex-typed domain (Marsh & Yeung, 1998). Third, mathematics is one of the domains of self-concept that is most strongly associated with learning gains (Marsh, Trautwein, et al., 2005; Marsh & Yeung, 1998) and with achievement-related choices (Marsh & Yeung, 1997b; Nagy et al., 2006, 2008).

Our investigation focused on development in mathematics self-concept from grades 7 to 12. In order to overcome limitations in the comparability of previous studies (i.e., Fredricks & Eccles, 2002; Jacobs et al., 2002; Watt, 2004), we selected data sets with self-concept measures based on similar theoretical conceptions, and used highly flexible growth curve models that did not rely on prespecified time contrasts.

Based on previous findings (Fredricks & Eccles, 2002; Jacobs et al., 2002; Watt, 2004), we expected to find a pattern of gender convergence for the United States. Because Watt (2004) used a measure that is theoretically different from self-concept, we did not formulate any specific predictions for gender convergence or gender differentiation in the Australian sample. Rather, we approached the question of whether self-concept follows the same gendered trends as perceived talent (Watt, 2000) as an open research question. Because gender differences in self-concept development have not previously been investigated in Germany, we did not make any predictions for gender convergence or gender differentiation in this context either.

Regarding the differences between the three countries, we expected to find stronger gender effects in favor of males in Germany than in Australia or the United States. Germany has been described as a cultural setting with highly differentiated gender norms (Williams & Best, 1990), which suggests larger gender differences (Lynn & Martin, 1997). Given the available research findings, however, we did not feel able to formulate a clear hypothesis on whether the differences would already be larger at the beginning of the observational time frame (i.e., in grade 7) or whether they would increase during the later school years.

Note that, although we drew on data from published studies, the analyses presented here are unique. First, for the United States, we used a different data set from that used by Fredricks and Eccles (2002) and Jacobs et al. (2002). Second, although we used the same Australian sample as Watt (2004), we used a measure of self-concept rather than perceived talent.

METHOD

Samples

Australia (Sydney). We used the same Australian sample as Watt (2004). Participants spanned grades 7–11 in a longitudinal cohort-sequential design with $N = 1{,}333$ students in three cohorts (cohort 1: $N = 428$, 45% females; cohort 2: $N = 436$, 44% females; cohort 3: $N = 459$, 43% females). Students were drawn from three upper middle class secondary schools in northern metropolitan Sydney, matched for socioeconomic status. The participants were predominantly of English-speaking background (73–82%), with Asians as the largest ethnic subgroup (12–22%). The first assessment took place at the end of 1995 and the last at the beginning of 1998. Approximately half of the participants in each cohort were present for all administrations, and approximately one fourth missed only one administration. More details on the sample composition can be found in Watt (2004).

Germany (four states). The German sample was derived from the "Learning Processes, Educational Careers, and Psychosocial Development in Adolescence and Young Adulthood" study (German acronym: BIJU), conducted at the Max Planck Institute for Human Development in Berlin, Germany (Baumert et al., 1996). Here, we concentrated on a subsample of the BIJU study, namely *Gymnasium* students ($N = 4{,}688$, 58.1% females). These students were assessed three times (in grades 7, 10, and 12). Data obtained from the low- (Haptschule) and middle-tier tracks (Realschule) were not considered in the present analyses because these two subsamples were assessed only twice, precluding application of growth curve techniques (the results are available from the first author upon request). The schools were randomly sampled in each of the four participating federal states. Almost all of the participating students had German citizenship. As is typical for Germany, over 95% of participants were Caucasian.

United States (Michigan). For the U.S. sample, data from $N = 2{,}378$ (52.7% female) adolescents were derived from the Michigan Study of Adolescent Life Transition (MSALT; e.g., Eccles, Lord, & Roeser, 1996). Eccles and her colleagues began collecting data in 10 school districts in southeastern Michigan in 1983, when participants were in grades 5 and 6. The majority of the participants in this study were White (87%) and from working or middle class families in small communities surrounding Detroit, Michigan. To date, 9 waves of data have been collected. The present study is based on waves 3, 5, and 6, when students were in grades 7, 10, and 12.

At this point, it is important to note the potential limitations of the present data sets. Because they were obtained from specific regional areas, the Australian and U.S. data sets are not representative for the countries as a whole. Hence, results derived from these data sources may be affected by regional peculiarities. The German data set is quite representative for the four states covered. Although the representativeness for Germany is unclear, the four federal states cover the west–east as well as the north–south divides and may thus be quite characteristic of Germany as a whole.

Self-Concept Instruments

Because the three studies were not designed to be strictly comparable, different instruments were used to measure mathematics self-concept in the three samples. Nevertheless, the measures used in Germany and Australia were based on a similar instrument (Marsh, 1992), and the items used in the MSALT study were developed by Eccles et al. (1983) (see also, Fredricks & Eccles, 2002; Jacobs et al., 2002) based on a similar theoretical rationale.

Australia (Sydney). In Australia, mathematics self-concept was measured by five items from Marsh's (1992) Self-Description Questionnaire with minor modifications (items rephrased as questions, responses ranging from 1 = *never* through 7 = *always*; sample item: "Do you get good marks in math?").

Germany. An established five-item German self-concept scale was used to measure domain-specific academic self-concept in mathematics (sample item: "Nobody's perfect, but I'm just not good at math") in the BIJU study. This scale has a very high latent correlation of $r = .97$ with a short form of Marsh's (1992) Self-Description Questionnaire (see Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2007). Students responded to each item on a 4-point 1 = *agree* to 4 = *disagree* scale.

United States (Michigan). In the MSALT sample, self-concept was measured by means of short but well-validated three-item scales (sample item: "How good at math are you?"), the psychometric properties of which have been reported elsewhere (Eccles et al., 1983). Students responded on a 7-point scale.

Statistical Procedures

Latent growth curve models (LGM). Self-concept trajectories were analyzed by means of LGM (Meredith & Tisak, 1990) on the construct levels using structural equation modeling (SEM) techniques. Analyses were conducted in Mplus 4.2 (Muthén & Muthén, 1998–2006).

A schematic representation of the model used for data with three time points is given in Figure 1. Instead of using the conventional approach that relies on scale scores, we analyzed the data on the item level. To this end, we first formulated occasion-specific measurement models linking the observed item responses to the underlying time-specific constructs. The measurement part was constrained to be measurement invariant across time points (i.e., invariant factor loadings and measurement intercepts). Furthermore, we allowed for correlated item uniquenesses across time to account for the effects of specific factors not accounted for by the underlying common factors. Because measurement invariance is crucial for the proposed models, we investigated this issue very carefully. We found clear support for strong measurement invariance (Meredith, 1993), meaning that factor loadings and measurement intercepts could be constrained to be invariant across time points in all samples considered. Establishing strong invariance is a prerequisite for interpreting the results of our LGM, because these imply a specific covariance structure of time-specific self-concept factors (requiring invariant factor loadings) in addition to a specific mean structure of the factors (requiring invariant measurement intercepts). Results are available upon request from the first author.

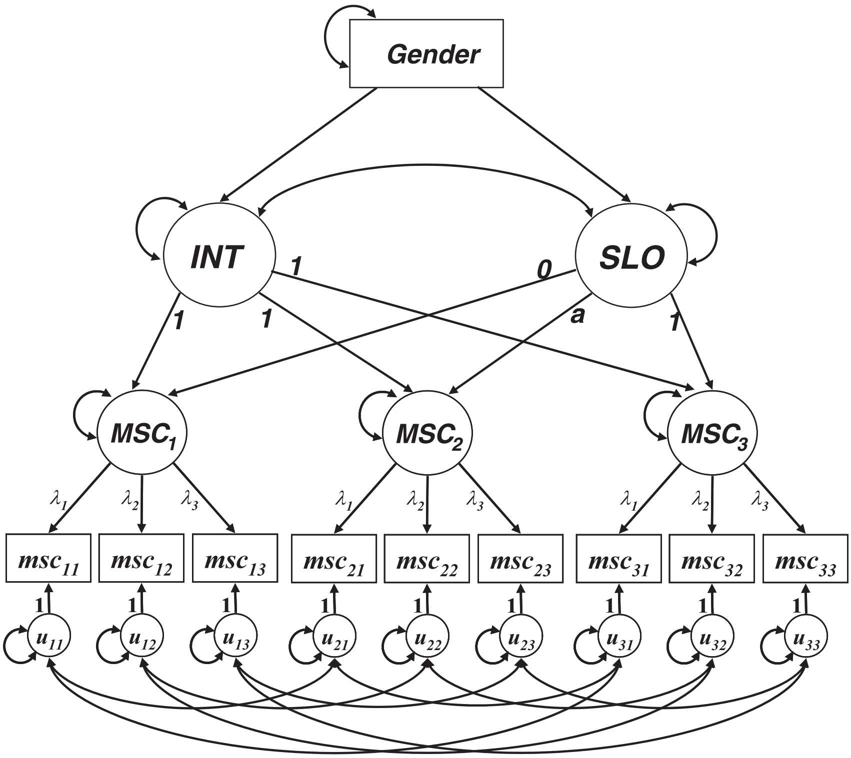

FIGURE 1 Schematic path diagram of the latent growth curve model used. Circles represent latent variables; squares stand for observed variables. Directed arrows stand for regression coefficients and factor loadings, bidirectional arrows connecting two variables stand for covariances/correlations, and bidirectional arrows attached to a single variable stand for variance terms.

The occasion-specific factors were used to define two characteristics that describe individuals' self-concept growth: one factor representing initial individual differences at the first occasion of measurement in grade 7 (initial level; INT in Figure 1), and one factor representing the amount of individual change (slope; SLO in Figure 1). The initial-level factor was identified by fixing all loadings to unity. In order to identify the change factor, we fixed its loading in grades 7 to 0 and its loading in grades 10 to 1. Thus, the mean of this factor represents the mean change between grades 7 and 10. The remaining loadings were not fixed to any prespecified value, but were directly estimated from the empirical data (Meredith & Tisak, 1990). As a consequence, the model represented in Figure 1 can accommodate any form of nonlinear change.

Finally, we included gender (0 = females; 1 = males) as a time-invariant covariate. Effects on the initial-level factor represent gender differences at the initial assessment, and effects on the change factor represent gender differences in the rate of change. Positive effects indicate higher values in males and intercepts of dependent variables reflect mean values for females.

Model fit was judged on conventional criteria: the χ^2-test, the CFI, and the RMSEA. The χ^2-test is a classic goodness-of-fit index that is routinely reported in empirical studies, despite some well-known limitations. The χ^2 measure tests the null hypothesis of zero deviations between model-generated data and observed data. A specified model is said to be falsified if the χ^2-test is statistically significant. However, it is well known (e.g., Brannick, 1995; Kelloway, 1995) that the test rejects correctly specified models far too often, especially when sample sizes are large (e.g., Hoyle, 1995). Hence, in the case of large-scale data—as used here—the χ^2 goodness-of-fit test is virtually always statistically significant. Because of the fundamental problems associated with the χ^2 statistic, we judged model fit mainly on basis of the practical fit indexes CFI and RMSEA. Several guidelines for cut-off values have been suggested. CFI values $>.90$ are typically taken to reflect an acceptable fit to the data (Marsh, Hau, & Grayson, 2005). RMSEA values $<.05$ and .08 are taken to reflect a close fit and a reasonable fit, respectively; values between .08 and .10 reflect a mediocre fit, and values $>.10$ are generally unacceptable (Browne & Cudeck, 1993). In order to be better able to judge model fit, we also report the RMSEA test of close fit, which gives the probability that the fit index is RMSEA $\leq .05$.

Missing data. Almost all longitudinal studies must contend with the problem of missing data. There were different reasons for missing data in the three data sets used in this investigation. The main reason for missing data in the Australian sample was the multicohort design implemented. As a consequence of this design, the number of participants increased on some occasions (i.e., the occasions on which more then one cohort was assessed; see Table 1). The main reason for missing data in the German sample was that some schools refused to participate in later assessments. In order to counterbalance these losses, the German samples were refreshed by drawing new schools. This design resulted in a substantial amount of missing data at each occasion (see Table 2). Finally, the main reason for missing data in the U.S. sample was sample attrition.

It has been shown that popular approaches to missing data, such as listwise and pairwise deletion or mean substitution of data, might bias results (Allison, 2001). Likelihood-based estimation procedures, such as full information maximum likelihood estimation (FIML; Arbuckle, 1996) or the multiple imputation of missing data, are highly preferable to most other approaches to this problem. In order to make full use of all available information we therefore used the FIML estimator in all analyses reported below and computed model parameters from all available data points.

TABLE 1
Descriptive Statistics for the Mathematics Self-Concept Scales in the Australian (Sydney) Sample

	N (N Missing)	*M*	*SD*	α	*Correlation With Gender*
Grade 7	744 (589)	4.45	1.23	.87	.11*
Grade 8	782 (551)	4.33	1.20	.88	.14*
Grade 9	1,172 (161)	4.35	1.11	.83	.13*
Grade 10	761 (572)	4.23	1.05	.82	.16*
Grade 11	356 (977)	4.24	0.93	.78	.17*

*$p < .01$.

Hierarchical data. The present data are typical for educational research, with student samples nested within schools. When data are clustered in this way, the assumption of independence of observations, crucial to all standard statistical procedures, is likely to be violated (Bryk & Raudenbush, 1992). However, intraclass correlation (ICC) analyses showed that the clustering effect was very low in all data sets (all ICCs < .05), indicating that no correction of standard errors was necessary.

RESULTS

Australia (Sydney)

Descriptive statistics for the Sydney area sample are summarized in Table 1. As shown, the self-concept scales had good internal consistencies on each measurement occasion. The raw data indicate a small decreasing trend in mean mathematics self-concept scores accompanied by decreasing variances. Furthermore, the mathematics self-concept scales were significantly correlated with gender, although the magnitude of the association was rather small.

In order to investigate change in mathematics self-concept, we applied an LGM as represented in Figure 1. In the first step, we estimated a model without gender effects. Model fit was satisfactory, $\chi^2(214, N = 1{,}333) = 961.5$,

TABLE 2
Descriptive Statistics for the Mathematics Self-Concept Scales in the German Sample (BIJU)

	N (N Missing)	*M*	*SD*	α	*Correlation With Gender*
Grade 7	3,533 (1,155)	2.98	.75	.91	.25*
Grade 10	2,685 (2,003)	2.86	.75	.87	.27*
Grade 12	1,762 (2,926)	2.77	.76	.88	.24*

*$p < .01$.

$p < .01$; CFI = .927; RMSEA = .051, $p = .27$. It needs to be reiterated that high and statistically significant χ^2 values—despite indicating "bad" fit—are to be expected in large samples such as the present one. The mean initial level was estimated to be 4.46 ($SE = .04$, $p < .01$) and the mean rate of change to be $-.14$ ($SE = .04$, $p < .01$), reflecting a decline in self-concept. Both factors, initial level and slope, had significant variance terms (level: 1.04 [$SE = .08$], slope: .46 [$SE = .10$], both ps $< .01$), meaning that there were reliable individual differences in intraindividual change.

In the next step, we extended the growth curve model to incorporate the time-invariant covariate gender. Model fit was again good, $\chi^2(237, N = 1{,}333) = 982.9$, $p < .01$; CFI = .927; RMSEA = .048, $p = .85$, indicating that the relation between gender and the time-specific manifestations of mathematics self-concepts could be conceived as being transmitted by the underlying factors. Boys had a higher mathematics self-concept than girls in grade 7 ($b = .31$, $SE = .06$, $p < .01$), but there was no gender difference in the amount of change ($b = -.02$, $SE = .06$, $p = .36$). Overall, this pattern of results indicates stable gender differences in mathematics self-concept that neither diminish nor increase over the course of secondary education in Australia. Figure 2 illustrates the mean self-concept trajectories for the total sample and by gender. As shown, the trajectories were nonlinear and appear to decline slightly more in senior grades.

Germany

Table 2 presents the descriptive statistics for the BIJU sample. Scale reliabilities were good for all occasions. Mean scores indicate a decreasing trend for mathematics self-concept. Correlations with gender suggest that gender differences remained stable across time.

In order to analyze self-concept change, we began with an unconditional LGM. This model achieved a good fit, $\chi^2(88, N = 4{,}688) = 914.4$, $p < .01$; CFI = .966; RMSEA = .045, $p = 1.00$. The German *Gymnasium* students had an initial self-concept level of 2.98 ($SE = .01$, $p < .01$) and a negative mean rate of change of $-.16$ ($SE = .02$, $p < .01$). Variance estimates were significant for both factors (level: .36 [$SE = .03$], slope: .13 [$SE = .02$], both ps $< .01$) documenting individual differences in self-concept change.

In the next step, we extended the unconditional model by integrating the time-invariant covariate gender. Model fit was good, $\chi^2(207, N = 8493) = 954.1$, $p < .01$; CFI = .966; RMSEA = .042, $p = 1.00$. Significant gender differences were found for the mean initial level of mathematics self-concept ($b = .37$, $SE = .02$, $p < .01$). However, there was no support for gender differences in self-concept change ($b = .01$, $SE = .02$, $p = .48$). Hence, the findings did not indicate increasing or diminishing gender differences in mathematical self-concept in the German *Gymnasium*.

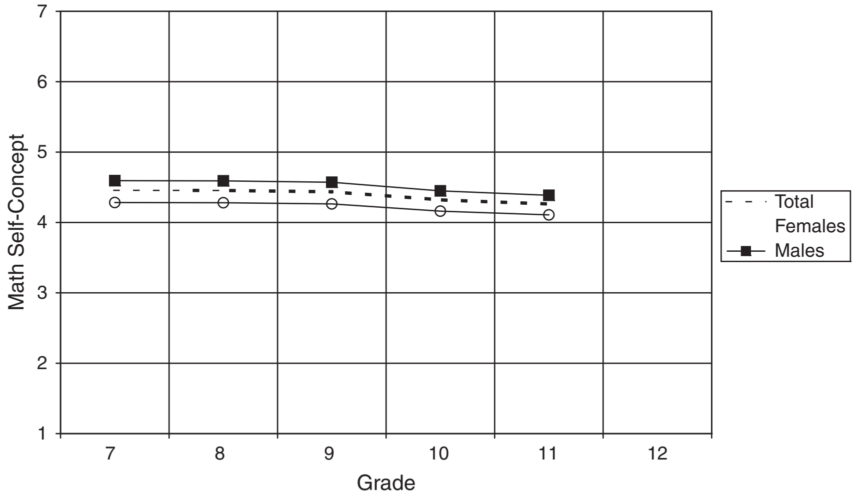

FIGURE 2 Developmental trends for mathematics self-concept in the Australian (Sydney) sample. Mean trends for the total sample and by gender.

In sum, the analyses of the German data document decreasing trends in mathematics self-concept. The results also provide evidence for gender differences in mathematics self-concept, but not for relations between students' gender and the amount of change in self-concept. This pattern of results is illustrated in Figure 3.

United States (Michigan)

Table 3 contains the descriptive results for the MSALT sample. The scales exhibited satisfactory reliabilities, and there was a clear decreasing trend in mathematics self-concept. Correlations with gender were consistent but rather low.

In the next step, we fitted the unconditional LGM to the raw data. Model fit was good, $\chi^2(23,\ N=2{,}443)=132.2,\ p<.01$; CFI = .986; RMSEA = .045, $p=.87$. The mean initial level of mathematics self-concept was estimated to be 5.03 ($SE=.03,\ p<.01$) in grade 7, and there was a significant negative trend from grades 7 to 12. The mean rate of change in self-concept scores was estimated to be $-.43$ ($SE=.04,\ p<.01$). Furthermore, as indicated by significant variances for the initial level (.39, $SE=.10,\ p<.01$) and for the slope factor (.15, $SE=.06,\ p<.05$), there were reliable individual differences in intraindividual change.

To examine the impact of gender on self-concept development, we extended the LGM to incorporate this time-invariant covariate. Model fit was again good, $\chi^2(30,\ N=2{,}443)=143.1,\ p<.01$; CFI = .985; RMSEA = .039, $p=.99$. Results showed that males and females differed in their mathematics

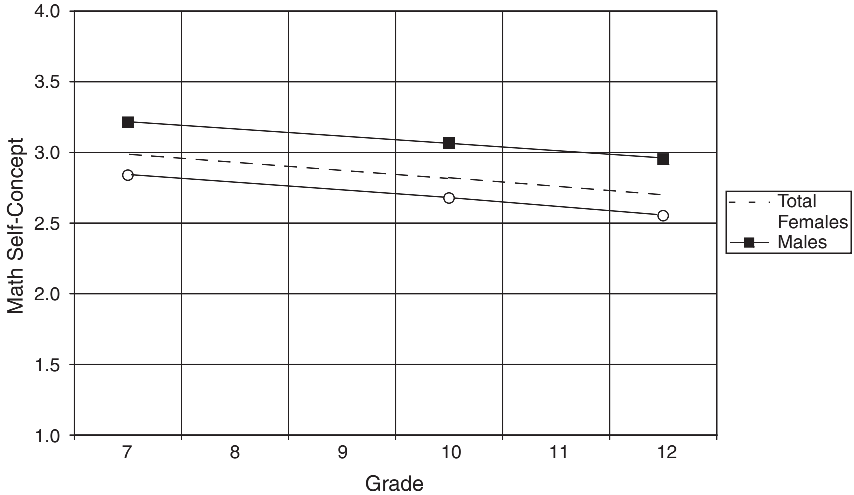

FIGURE 3 Developmental trends for mathematics self-concept in the German sample (four states). Mean trends for the total sample and by gender.

self-concept in grade 7, with males scoring higher ($b = .21$, $SE = .05$, $p < .01$), but not in their rates of change ($b = .01$, $SE = .04$, *ns*). Females had an initial level of 4.93 ($SE = .03$, $p < .01$) and a mean rate of change of $-.43$ ($SE = .04$, $p < .01$).

Figure 4 illustrates the mean self-concept trajectories for the total sample and by gender. As shown, the trajectories appear to be linear across the three time points under consideration.

Comparison of Gender Effects Across Groups

In the previous sections, we have documented gender differences in mathematics self-concept and its development. In this section, we compare the results determined for each data set, considering differences in the form of self-concept change and in the effect of gender on self-concept.

TABLE 3
Descriptive Statistics for the Mathematics Self-Concept Scales in the Michigan Study of Adolescent Life Transition (MSALT) Sample (Michigan)

	N (N missing)	*M*	*SD*	α	*Correlation With Gender*
Grade 7	2,266 (177)	5.04	1.21	.83	.08*
Grade 10	1,414 (1,029)	4.64	1.34	.86	.08*
Grade 12	1,280 (1,163)	4.39	1.35	.87	.09*

*$p < .01$.

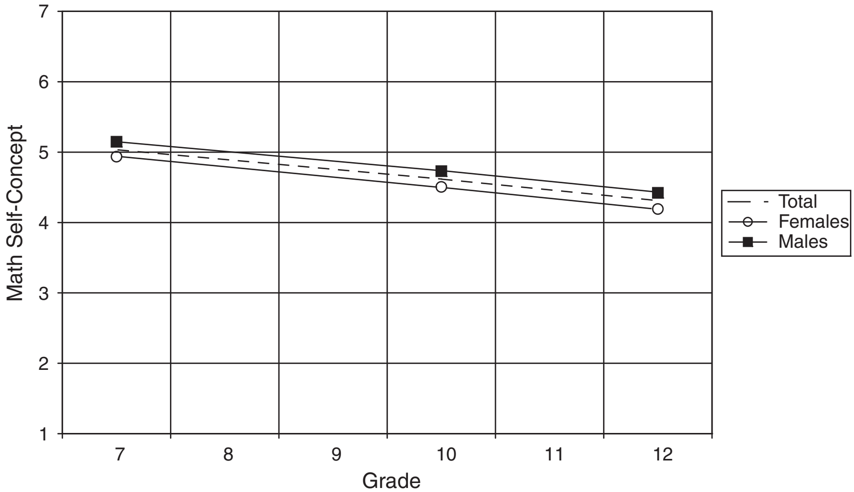

FIGURE 4 Developmental trends for mathematics self-concept in the U.S. sample (Michigan; Michigan Study of Adolescent Life Transition). Mean trends for the total sample and by gender.

The comparison of the form of change is based on the freely estimated factor loadings of the time-specific self-concept factors on the slope factor. As stated above, the growth factor was identified by fixing its loading on the grade 10 self-concept to 1. The remaining loadings can thus be interpreted in reference to the change from grades 7 to 10. The loading for grade *X* reflects the change from grades 7 to *X* relative to the change between grades 7 and 10. For example, a loading of 0.8 for grade *X* means that the change that has occurred between grades 7 and *X* is 80% of the size of the change between grades 7 and 10. Because the loadings are not sensitive to differences in the metrics of the analyzed items, they can be compared across groups. Note that the potential for comparing the functional form of self-concept across settings is rather restricted in this study on account of the limited number of data points available. Our comparison of the form of change is thus intended to be exploratory and descriptive.

Figure 5 plots the loadings from the Australian (Sydney), German (four states), and U.S. (Michigan) samples. As shown, change in self-concept was nonlinear in Australia, with self-concept remaining more or less stable until grade 9 and then declining until grades 11 or 12 (see also Figure 2).

The form of change in the Michigan sample and in the German academic track appeared to be linear (see also Figures 3 and 4), but because data were available from only three time points (grades 7, 10, and 12), we cannot assume that this pattern generalizes across all grades. Based on the present data, it appears that self-concept change can be described by a linear function

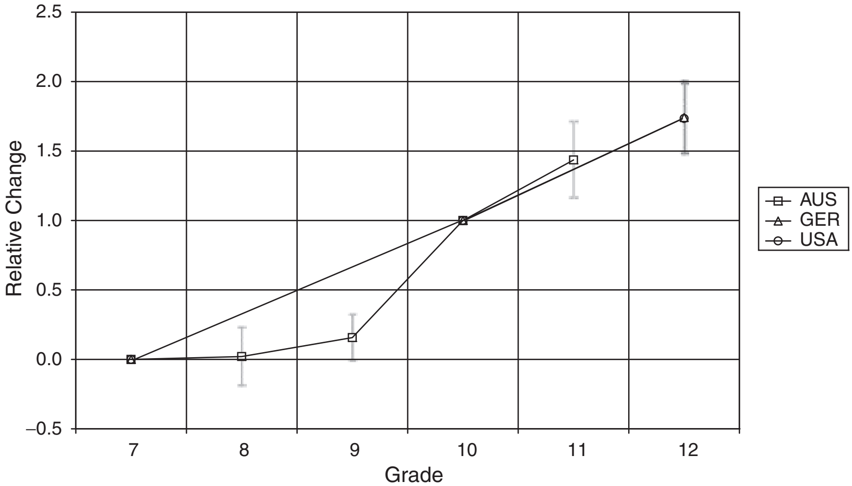

FIGURE 5 Plot of slope factor loadings determined for the Australian (Sydney), German (four states), and U.S. (Michigan) samples. Scale bars represent 95% confidence bands. Grades 7 and 10 loadings do not have confidence bars because these loadings were fixed.

across the grades under consideration in both settings. We tested whether self-concept change in the Australian (Sydney) sample was adequately described by a linear change function. This was not the case, however; imposing the corresponding constraints led to a statistically significant reduction in model fit: $\Delta\chi(3, N = 1{,}333) = 67.2$, $p < .01$.

The combined pattern of factor loadings—including Australia (Sydney)—suggests that self-concept change is linear in all settings under investigation between grades 10 and 12. Of course, additional data from grades 11 in the United States and Germany and from grade 12 in Australia may paint a different picture. Based on the present data, however, we cannot conclude that the form of self-concept development in senior grades differs across the three settings investigated.

We next investigated whether the gender effect on self-concept varies across settings. A certain complication occurred here because the self-concept measures used in the different primary investigations were not strictly comparable (i.e., different items with different response formats). We thus had to rely on standardized effects. Specifically, we report *Y*-standardized regression coefficients (e.g., Long, 1997) that give gender differences in standard deviation units. The metric is similar to Cohen's *d*, with the exception that it is applied to latent variables that are free of measurement error.

Table 4 presents *Y*-standardized effect sizes and their 95% confidence bands. As reported above, gender differences in slope were not significant in any group. Furthermore, given the overlap of confidence bands across groups, we concluded that these effects did not vary across settings. A rather different picture emerged for gender differences in initial levels, for which we already have documented consistent and significant effects. As Table 4 shows, the largest gender differences were found in the German sample. Because the confidence bands for this group did not overlap with those determined for Australia (Sydney) or the United Sates (Michigan), we concluded that the gender effect on self-concept is significantly larger in the four states of Germany examined in the present study. Given the high overlap of confidence bands between Sydney and Michigan, gender effects in these samples appear to be of similar magnitude.

DISCUSSION

This is the first paper to compare students' mathematics self-concept development in the secondary years in three distinct cultural settings: Australia (Sydney), the United States (Michigan), and Germany. The analyses presented drew on large student samples and overcame some of the shortcomings of previous attempts to compare the results of different primary investigations. First, the self-concept measures administered were based on a similar theoretical rationale. Second, the period of observation was comparable across settings. Third, we used a class of highly flexible models of change. Fourth, we extended previous research by examining self-concept change in Germany, a cultural context characterized by more highly differentiated gender norms (Williams & Best, 1990) and with a rather different school structure from those examined in previous studies, thus providing further insights into cultural differences in the self-concept development of males and females.

Given the substantial differences between the settings under consideration, the similarities in the patterns of results are surprising. First, a general decline in mathematics self-concept was evident in all three countries. There was no evidence for qualitative differences across settings in this pattern of

TABLE 4

Y-Standardized Gender Effects on Initial Levels (Grade 7) and Rates of Change (Grades 7–10) of Mathematics Self-Concept and 95% Confidence Bands (Standard Errors in Parentheses)

	Australia (Sydney)		*Germany (Four States)*		*United States (Michigan)*	
	Initial Level	*Slope*	*Initial Level*	*Slope*	*Initial Level*	*Slope*
β^y	.31 (.06)	– .04 (.09)	.62 (.04)	.04 (.06)	.34 (.03)	.03 (.11)
95% CI	.18; .43	– .21; .15	.54; .71	– .07; .15	.17; .52	– .19; .24

Note. Results for the Australian, U.S., and German (academic track) samples.

change. Second, consistent gender differences in mathematics self-concept were found in all settings. Males scored higher than females at the beginning of grade 7 in all contexts, but the magnitude of these gender differences varied. The largest differences were found for Germany, with smaller gender gaps in Australia (Sydney) and the United States (Michigan). Third, and most surprising, there was no evidence for gender differences in self-concept change (consistent with patterns for perceived talent; Watt, 2004). This pattern is inconsistent with both the *gender intensification* model and the *gender convergence* hypothesis.

Taken together, findings suggest that—at least for the cultural settings examined, the period from grades 7 to 12, and the mathematical domain—initial gender differences in self-concept persist unchanged. This does not imply that there are no gender differences in self-concept development per se. It seems more likely—at least for the mathematical domain—that such differences materialize before or/and after the period considered here. Indeed, the findings provided by Jacobs et al. (2002) suggest that the largest gender-related changes occur between grades 1 and 7, a period that was not included in the present study.

Given the absence of gender differences in change in mathematics self-concept, the consistent patterns of results across settings are the most interesting findings from the present research. In line with previous work, we found a general decline in self-concept across the secondary years. The decreasing trend in perceived mathematical competence is perplexing to the extent that students' objective domain-specific and general intellectual abilities clearly increase over the time period considered (e.g., Baltes, Staudinger, & Lindenberger, 1999). Students acquire a substantial amount of declarative and procedural knowledge about mathematical topics over this period, and their fluid intelligence (Cattell, 1963), which is highly related to mathematics achievement (Ackerman, 1996), also improves. Hence, the self-concept trajectories observed do not mirror intellectual development in adolescence.

Various explanations have been proposed for the pattern of decreasing self-concepts. Some researchers see it as a result of the psychological and physical changes occurring at puberty (e.g., Blyth, Simmons, & Carlton-Ford, 1983). Others consider it to result from differences in the classroom environment before and after the transition to secondary school, with a lack of fit between the high school environment and the needs of adolescents negatively impacting on their development (Eccles & Midgley, 1989, 1990). The latter perspective has been challenged by Baumert, Trautwein, and Nagy (2005), who have shown that students' self-perceptions decline even in domains to which they have not been exposed at school. Although the reasons for the observed decline remain unclear, the present results provide additional evidence that the phenomenon generalizes to different cultural settings.

Indeed, the present findings are congruent with the notion that self-concept decline is related to both contextual (i.e., school curriculum) and biological (i.e., puberty) factors. Results from LGMs applied to the Australian data document steeper declines in senior grades, a period in which curricular demands are increasing. The pattern of self-concept decline observed in the Australian (Sydney) sample is also in line with biological explanations: The steepest declines occurred between grades 9 and 10, when most students were between 14 and 16 years old, the time when the most dramatic pubertal changes are experienced. The pattern of results is thus in line with contextual and biological explanations. Additional data, including a larger number of repeated observations accompanied by fine-grained measures of contextual and biological factors and their change, are required before more definitive conclusions can be drawn.

The varying gender differences across groups under consideration also warrant discussion. As noted by others, such differences have often no realistic basis, as girls often outperform boys in mathematics (e.g., Marsh & Yeung, 1998; Wigfield et al., 1997). Nevertheless, some research suggests that gender differences in mathematics self-concept emerge as early as the beginning of primary school (Fredricks & Eccles, 2002; Jacobs et al., 2002). From this perspective, stereotypic gender roles seem to have an important impact early in individual developmental trajectories. Our research suggests that such stereotypes seem to be more influential in some countries (e.g., Germany) than in others. This finding is line with the higher gender norm differentiation reported for German society (Williams & Best, 1990). However, it has to be noted that the evidence for societal influences provided here is only indirect. Further cross-cultural comparative studies are needed in which gender norms are directly assessed and linked to students' self-concept development.

The varying gender differences across studies may also be due to differences in school structure—in the present investigation, this characteristic was confounded with cultural differences in gender norms. However, this explanation does not seem very likely. Indeed, we conducted additional analyses (not reported here) comparing students' self-concept development across the three German school tracks, and the results were very similar to those obtained in the *Gymnasium*. However, to gain a better sense of the factors operating, future research should analyze data sets that allow effects of cultural norms to be better separated from effects of the imposed school structure.

Limitations and Further Research

Given that the data used in the present article were taken from primary investigations that were not designed to be comparable, the analyses could be improved in a number of ways. First, confidence in results could be enhanced by ensuring that the same measures are used in all groups under comparison. Using the same instruments has the advantage of allowing raw effects that are not affected by possible differences in variance terms to be compared across cultures.

Furthermore, it would make it possible to investigate whether the constructs differ in the meaning attached to them in the cultures under consideration.

Second, longitudinal data could profitably be collected at more occasions of measurement. As we have shown, the pattern of self-concept change can be very complex. In order to detect more subtle differences in development of self-concept, data need to be gathered more frequently, ideally in the same grades in all groups under investigation.

Third, although the cultural settings and educational structures investigated are relatively different, Australia, the United States, and Germany also have much in common. For one thing, they are all characterized as Western individualistic societies. It would be interesting to extend the present research to Eastern collective cultures.

Fourth, the data sets used here are not representative for the countries as a whole. Hence, results derived from these samples may be affected by regional peculiarities. Given that we did not find support for gendered patterns of self-concept development in any of the samples, however, we consider it unlikely that these consistent results are solely attributable to regional characteristics. Replications based on more representative samples are needed to provide more definitive answers.

Conclusion

Mathematics self-concept is a strong predictor of school achievement and of educational and occupational choices. Furthermore, gender differences in mathematics self-concept account for a good deal of the gender effect on activity choices. From this perspective, it is important to gain a deeper understanding of self-concept development in males and females. As shown in the present article, the development of mathematics self-concept appears to be highly similar across cultural settings. No qualitative differences were found in the form of self-concept change, and gender relations with self-concept development were absent in all investigated settings. Taken together, the present findings for grades 7–12 do not support the predictions of either gender intensification theory or the gender convergence hypothesis. On the contrary, the period under investigation seems to be characterized by stable gender differences across cultures.

APPENDIX A

This appendix gives a short description of differences in the educational systems of the countries under consideration, namely, Australia, Germany, and the United States.

In Australia, secondary school serves grades 7–12. The mathematics syllabus for grades 7 and 8 focuses largely on consolidating material learned in primary grades 3–6. In grades 9 and 10, students are streamed into advanced, intermediate, and standard mathematics classes based on their demonstrated ability up to that point. In grades 11 and 12, which lead up to the Higher School Certificate, a major external examination supplemented by within-school assessment, students elect which subjects to study. Although mathematics is not compulsory, the overwhelming majority of students opt to take the subject, which is generally perceived to be important and is a requirement for many tertiary courses. Students preparing for the Higher School Certificate also select within each academic domain which difficulty level they wish to undertake.

In Germany, students are tracked into the different schools of the three-tiered secondary system from as early as age 10 (i.e., grade 5). The three main secondary school types (*Hauptschule*, *Realschule*, and *Gymnasium*) differ greatly in the intensity of the curriculum, and there is little tracking within each school type. The exception is the most academically competitive track, the *Gymnasium*, in which course selection is an integral part of the last 2 years of schooling (grades 12 and 13 in most states).

The *Hauptschule*, *Realschule*, and *Gymnasium* also differ in terms of the years of schooling provided and the school leaving certificate awarded. *Hauptschule* students graduate after grades 9 or 10, *Realschule* students after grade 10, and *Gymnasium* students after grades 12 or 13. The school leaving certificates issued by the *Hauptschule* and *Realschule* qualify students to enter vocational education and training; the level of vocational education for which they are eligible depends on the school type attended. Students who graduate from *Gymnasium* are awarded the *Abitur*, which qualifies them for college entry. Streaming in the three-tiered German school systems is quite strict; students assigned to one of the three school types tend not to change tracks before graduation (Schnabel et al., 2002). The tracks of the three-tier system can thus be characterized as distinct developmental niches (Baumert, Köller, & Schnabel, 2000).

In the United States, all secondary students attend high school, where within-school tracking usually starts at grades 8 or 9. However, there is no official age or grade level by which students and their families must decide on future educational pathways and the possibility of postsecondary education. It is possible to identify students who are on a "college-bound" track by the courses they select ("college preparatory"), although this distinction is less clear in the United States than it is in Germany. In addition, students in the United States have a great deal of freedom to choose among classes. Assuming they have taken the necessary prerequisite courses (i.e., have been on the "right" track), students are relatively free to choose between higher and lower level classes in each subject.

REFERENCES

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, knowledge. *Intelligence*, *22*, 227–257.

Alexander, K., & Entwisle, D. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development*, *53*. Serial No. 218.
Allison, P. D. (2001). *Missing data (Sage university papers series on quantitative applications in the social sciences, 07-136)*. Thousand Oaks, CA: Sage.
Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.
Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology, 50*, 471–507.
Baumert, J., Köller, O., & Schnabel, K.-U. (2000). Schulformen als differentielle Entwicklungsmilieus — Eine ungehörige Fragestellung? [School types as differential developmental milieus — An improper question?]. In M. Demmer (Ed.), *Messung sozialer Motivation: Eine Kontroverse* (pp. 28–69). Frankfurt am Main, Germany: GEW.
Baumert, J., Roeder, P. M., Gruehn, S., Heyn, S., Köller, O., Rimmele, R., et al. (1996). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter* (BIJU) [Learning processes, educational careers, and psychosocial development in adolescence (BIJU)]. In K.-P. Treumann, G. Neubauer, R. Moeller, & J. Abel (Eds.), *Methoden und Anwendungen empirischer pädagogischer Forschung* (pp. 170–180). Münster, Germany: Waxmann.
Baumert, J., Trautwein, U., & Nagy, G. (2005, August). *Self-concept and interest trajectories in secondary education: Differentiation and selective optimization*. Paper presented at the 11th Biennial Conference of the European Association for Research on Learning and Instruction, Nicosia, Cyprus.
Blyth, D. A., Simmons, R. G., & Carlton-Ford, S. (1983). The adjustment of early adolescents to school transitions. *Journal of Early Adolescence*, *3*, 105–120.
Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*, 201–213.
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage Publications.
Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
Cattell, R. B. (1963). Theory of fluid and crystalized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.
Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.
Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*, 135–172.
Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco: W. H. Freeman.
Eccles, J. S., Adler, T. F., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*, 26–43.
Eccles, J. S., Lord, S. E., & Roeser, R. W. (1996). Round holes, square pegs, rocky roads, and sore feet: The impact of stage/environment fit on young adolescents' experiences in schools and families. In S. L. Toth & D. Cicchetti (Eds.), *Adolescence: Opportunities and challenges* (Vol. 7, pp. 49–93). New York: University of Rochester Press.
Eccles, J. S., & Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for young adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 139–181). New York: Academic Press.
Eccles, J. S., & Midgley, C. (1990). Changes in academic motivation and self-perceptions during early adolescence. In R. Montemayor, G. R. Adams, & T. P. Gullotta (Eds.), *Advances in*

adolescent development: From childhood to adolescence (Vol. 2, pp. 134–155). Newbury Park, CA: Sage.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*, 215–225.

Eccles, J. S., Wigfield, A., Harold, R., & Blumfeld, P. (1993). Age and gender differences in children's achievement self-perceptions during the elementary school years. *Child Development*, *64*, 830–847.

Feather, N. T. (1988). Values, valences, and course enrollment: Testing the role of personal values within an expectancy-value framework. *Journal of Educational Psychology, 80*, 381–391.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519–533.

Frome, P., & Eccles, J. A. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 2*, 435–452.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. C. Petersen (Eds.), *Girls at puberty* (pp. 201–228). New York: Plenum.

Hoyle, R. H. (1995). *Structural equation modeling*. Thousand Oaks, CA: Sage.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior, 16*, 215–224.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 countries. *Journal of Social Psychology, 137*, 369–373.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81*, 417–430.

Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept: A test manual and a research monograph*. MacArthur, NSW, Australia: University of Western Sydney, Faculty of Education.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*, 341–360.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McCardle (Eds.), *Contemporary psychometrics: A Festschrift to Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Lawrence Erlbaum Associates.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2007). Big fish little pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*, 631–669.

Marsh, H. W., & Yeung, A. S. (1997a). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, *89*, 41–54.

Marsh, H. W., & Yeung, A. S. (1997b). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal*, *34*, 691–720.

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*, 705–738.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthén & Muthén.

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

Nagy, G., Trautwein, U., Köller, O., Baumert, J., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, *12*, 323–345.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis*. Mahway, NJ: Lawrence Erlbaum Associates.

Schnabel, K., Alfeld, C., Eccles, J. S., Köller, O., & Baumert, J. (2002). Parental influence on students' educational choices in the U.S.A. and Germany: Different ramifications — same effect? *Journal of Vocational Behavior, 60*, 178–198.

Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*, 334–349.

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98*, 438–456.

Watt, H. M. G. (2000). Measuring attitudinal change in mathematics and English over the first year of junior high school: A multidimensional analysis. *Journal of Experimental Education*, *68*, 331–361.

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*, 1556–1574.

Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., & Midgley, C. (1991). Transitions at early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552–565.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., & Blumenfeld, P. C. (1997). Changes in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology, 89*, 451–469.

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study.* Newbury Park, CA: Sage.