CrossMark

Academic self-concept and achievement in Polish primary schools: cross-lagged modelling and gender-specific effects

Paweł Grygiel[1] · Michał Modzelewski[1] · Jolanta Pisarek[1]

Abstract This study reports relationships between general academic self-concept and achievement in grade 3 and grade 5. Gender-specific effects were investigated using a longitudinal, two-cycle, 3-year autoregressive cross-lagged panel design in a large, representative sample of Polish primary school pupils ($N = 4226$). Analysis revealed (a) reciprocal relations between general academic self-concept and achievement over time but the influence of prior achievement on self-concept was stronger; (b) on average, levels of both constructs declined over time; (c) gender differences were not observed in longitudinal relations (i.e. cross-lagged, autoregressive and intra-wave correlations); (d) girls demonstrated higher mean levels of academic achievement at both grades; and (e) average level of general academic self-concept was not gender differentiated in grade 3 but decreased more for girls. These results are discussed in the light of the theoretical and practical implications.

Keywords Academic self-concept · Academic achievement · Gender effects · Longitudinal study · Primary school

Introduction

Structure of academic self-concept and its operationalization

According to the model proposed by Shavelson et al. (1976), academic self-concept is one aspect of the hierarchical and multi-faceted self-concept construct. This complex structure has its origins in individual experience and is subject to an influence from the environment and significant others but, in turn, also influences the way an individual acts. Self-concept is subject to developmental differentiation, with increasing separation of specific aspects with age (Marsh 1989; Marsh and Ayotte 2003; Marsh and Shavelson 1985). Structure of academic self-concept

---

✉ Paweł Grygiel
  p.grygiel@ibe.edu.pl

[1] Educational Research Institute (IBE), ul. Górczewska 8, 01-180 Warszawa, Poland

has been found to be internally differentiated. Two uncorrelated facets—mathematical and verbal academic self-concept—are required in order to reflect properly relationships between the multiple subject-specific self-concepts and the non-specific school self-concept (Marsh 1986; Marsh et al. 1988). Of the many possible operational definitions for academic self-concept, the one that uses general school self-concept scale seems recommended when more specific measures are not available or not applicable to the curriculum without clear subject division (Marsh et al. 1988; Marsh and Yeung 1997; Szumski and Karwowski 2015).

Implications for a causal relationship between academic self-concept and achievement

Academic self-concept is demonstrably the most important predictor for many types of behaviour and belief with long-term consequences on individual future prospects and labour-market position, i.e. motivation to learn (Guay et al. 2010; Marsh et al. 2006), educational and occupational aspirations (Marsh 1991; Nagengast and Marsh 2012), coursework selection (Marsh and Yeung 1997), career intentions and choices (Guay et al. 2004), psychological well-being (Craven and Marsh 2008), life satisfaction (Chang et al. 2003) and mental health (Marsh et al. 2004). Academic self-esteem is also closely tied to academic achievement (e.g. Falchikov and Boud 1989; Hansford and Hattie 1982; Mabe and West 1982; Zell and Krizan 2014). This positive correlation does not necessarily imply causality (Marsh 1990b). Therefore, the causal ordering between these constructs poses a major research question (Byrne 1984, 1986) in the study of academic self-concept. This critical aspect has not only theoretical but also important practical implications, since intervention to enhance academic achievement through enhancing self-concept (O'Mara et al. 2006a, b) would be redundant if academic self-concept could not be proved to benefit academic achievement (O'Mara and Marsh 2006).

Self-enhancement, skill development and reciprocal effect models

A review of literature reveals three principle theoretical models for causal ordering between both constructs: self-enhancement (Calsyn and Kenny 1977), skill development (Calsyn and Kenny 1977) and the reciprocal effects model of Marsh (1990b). The self-enhancement model suggests that self-concept is the primary determinant of academic achievement Support for this model would justify interventions focused on self-concept (Marsh and Martin 2011). By contrast, the skill development model implies causative influence on academic achievement by academic self-concept. Support for this model would provide stronger justification for the claim that development of academic skills is the best means to enhance self-concept (Marsh et al. 2012).

The reciprocal effects model, in turn, assumes that academic self-concept and academic achievement are reciprocally related and mutually reinforcing, i.e. a cross-lagged relationship exists between academic self-belief and academic achievement. According to this model, improved academic self-concept would lead to better academic achievement and improved academic achievement, to better academic self-concept (Marsh and Craven 2006). In consequence, based on this model, effective intervention should simultaneously impact academic achievement and academic self-concept (Marsh and Martin 2011).

A substantial body of research, reviewed by Marsh and Craven (2006), Marsh (2007) or Rosen (2010) with supporting meta-analysis by Valentine et al. (2004), found—with only a few notable exceptions—consistent support for reciprocal effects between academic self-belief

and achievement, although this effect was more commonly observed in adolescents than pre-adolescents (Marsh et al. 2007).

Age and changes in academic self-concept

Academic self-concept is subject to age-related dynamics and declines from a young age through adolescence (e.g. Eccles 1993; Eccles et al. 1993; Fredricks and Eccles 2002; Jacobs et al. 2002; Marsh 1985, 1989, 1990a; Marsh et al. 1984; Marsh et al. 1991; Marsh et al. 2005a, b; Nagy et al. 2010; Wigfield et al. 1997). Studies by Eccles et al. (1993), Marsh (1989) and Marsh et al. (1984, 1991) indicate that in the first grade of primary school, students are not only already able to evaluate their own competence but also to some extent to differentiate between the knowledge-dependent skills in various areas. There are, however, studies that suggest later development of this ability (Craven et al. 2000; Marsh 1990a; Marsh et al. 1984, 1991).

It is relevant that early child self-concept is characterised by unrealistic overestimation of personal abilities (Dweck 2002). Their perception is only marginally coming from external criteria, such as grades or test scores (Craven et al. 2000; Marsh 1985; Wigfield and Karpathian 1991). After a few years of schooling, self-concept becomes differentiated, increasingly abstract, includes more psychological descriptors and is based more strongly on social comparison with other children (Anderman and Maehr 1994). It is also more systematically related to external outcomes, i.e. parental or teacher feedback (Herbert and Stipek 2005; Marsh and Gouvernet 1989; Wigfield et al. 1997).

Development may prevent researchers from observing a reciprocal prospective relationship between achievement and academic self-concept in the younger age groups, since academic self-esteem may depend on prior achievement and not vice versa (Chapman and Tunmer 1997; Helmke 1995; Newman 1984; Skaalvik 1997; Skaalvik and Valås 1999). These results may then suggest a skill development model for younger children and a reciprocal effects model for older age groups (Fraine et al. 2007). Otherwise, some research (Guay et al. 2003; Muijs 1997; Quirk et al. 2009) even supported the reciprocal effects model for primary school pupils. In summary, only an unclear relation between these constructs for primary school children is evidenced from extensive review of the literature.

Gender effect in academic self-concept during pre-adolescence

Numerous studies report gender differences in level of academic self-concept (see Jansen et al. 2014, for review). A factor that significantly influences their nature is—in line with the multidimensional model of academic self-concept—the specific subject to which self-concept refers. In general, girls show greater self-concept verbally but less with reference to mathematics, biology and physics. It is particularly significant that girls characteristically demonstrate a greater level of general academic self-concept (Marsh 1985, 1989, 1990a; Marsh et al. 1984).

Empirical studies do not offer a uniform resolution of how gender predetermines changes in self-concept. Some indicate similar trajectories (lack of differences) for both sexes, and others point to increasing differences (for review, see Fraine et al. 2007). For general academic self-concept, findings from empirical research have also produced a somewhat equivocal pattern. Some studies conducted in primary schools (Marsh 1989; Marsh et al. 1984) reported the absence of impact of gender on changes in academic self-concept, whereas others described girls' greater age-related decline in general academic self-concept (Marsh 1985).

Gender gap in school grades

Recent meta-analysis of gender differences in teacher-assigned grades demonstrated a small but significant female advantage (Fischer et al. 2013; Richardson et al. 2012; Voyer and Voyer 2014). Many studies showed that throughout elementary, middle and high school, girls achieved higher grades than boys in all major subjects, even though their scores were lower or equal in tests, especially maths and science (for research reviews, see Duckworth and Seligman 2006; Ekstrom 1994; Kling et al. 2013). These conclusions were also confirmed by some meta-analyses (Else-Quest et al. 2010; Hyde et al. 1990; Lindberg et al. 2010).

Gender-specific differences in educational achievements in favour of girls are observed and often described, together with the gender gap in college attendance and behaviours, as the 'boy crisis' in many international studies (Voyer and Voyer 2014), but for fourth grade students, the issue is more complex (Mullis et al. 2012; Perie et al. 2005). Results from the PIRLS study showed that girls outperformed boys in almost all countries. Little reduction in the reading achievement gender gap has been noticeable over the decade. However, the TIMSS study demonstrated no significant gender difference in mathematics or science achievement in the nearly half the countries it covered. Small differences in favour of boys were found in most remaining countries.

Furthermore, the largest nationally representative assessment of American student achievements (NEAP) showed that girls outperformed boys in reading at all levels, as assessed by the study. However, gaps between girls and boys were smaller in fourth grade and increased in the eighth and 12th grades. Overall, girls outperformed boys in reading and writing by greater margins than boys' outperformance of girls in mathematics, science and geography. This gender gap at ages 9 and 13 have not much changed since 1971. In regards to gender differences in school achievement measured by teacher-assigned marks, meta-analysis (Voyer and Voyer 2014) demonstrated female advantage in all fields of study, but effect sizes were generally small. Moreover, female outperformance in terms of marks seemed stable across the years in data ranging from 1914 to 2011. Hence, discussion of the 'boy crisis' should concentrate more on the dynamics describing the decline in boys' performance and improvement of girls' performance (Mead 2006; Vail 2006).

This gender gap in school marks may have several plausible explanations (for reviews, see Burusic et al. 2012; Hadjar et al. 2014; Voyer and Voyer 2014). It is, of course, possible that grades—in line with the female underprediction effect theory (Shibley Hyde and Kling 2001)—may reflect additional cognitive abilities that are not captured in standardised tests (Conger and Long 2010), but research suggested that girls' generally better behaviour in the classroom and differential teacher expectations lead to their higher grades (Kimball 1989; Mullola et al. 2012; Spilt et al. 2012).

Review of the limited research showed that gender did not influence decline in academic performance, as measured by the GPA (e.g. Kling et al. 2013; Wang and Eccles 2012). In studies analysing marks in English and mathematics separately, the results were more mixed. For mathematics, the difference in grades to the girls' advantage increased with time (Downey and Vogt Yuan 2005; Shapka 2009). Otherwise, for English grades, no gender effect was observed on change (Downey and Vogt Yuan 2005). Owing to the dearth of available research and methodological inconsistencies (including age of population, incomparable sample sizes and the various approaches to measuring school grades), the summary presented above cannot be considered conclusive.

Theoretical summary

Conclusions from studies on the relationship between academic self-concept and academic achievement in primary school children were—as has been shown—varied. Some offered support for the skill development model and others for the reciprocal effects model. Such lack of consistency supports the replication of research. Investigation should not only use satisfactory statistical models (Marsh and Martin 2011) but should also test the underlying premises, related to time and group invariance (Little 2013; Little et al. 2007).

It is noteworthy that the majority of analysis in this respect did not examine mean level changes in constructs between subsequent study waves, most often being limited to determination of auto-predictive and cross-lagged relations. In effect, little is known as to whether gender differences in levels of academic achievement stem from differentiation in the dynamics of change of self-concept or vice versa, which would be helpful to recommendation of gender differentiated intervention.

Another limitation of previous studies on causal relation between academic self-concept and school achievement is reliance primarily on research from Western countries, particularly English-speaking students in Australia, Canada and the USA (Marsh et al. 2002). This overrepresentation may neglect potential cross-cultural differences, especially between individualistic and collectivistic societies (Marsh and Köller 2004). So, the reciprocal effects model requires confirmation in cross-cultural settings to prove its generalizability (Marsh et al. 2015; Marsh and Craven 2006; Wang 2006).

In the present study, we explored the relationship between academic self-concept and academic achievement and its potential differentiation with respect to gender from data in two waves of a large, longitudinal study conducted on a representative Polish primary school sample. Data was collected from children in grades 3 and 5 (aged 10 and 12). Following Marsh and Martin (2011), our analysis used multiple indicator structural equation models (SEM) (Little 2013) accounting for measurement error and so was expected to yield more accurate parameter estimates (Marsh et al. 1998).

Problems and hypotheses

The study was guided by a series of research questions, supported by hypotheses drawn from reviews of earlier work, as described above.

1. What is the relationship between prior academic self-concept and later academic achievement ?
2. How do mean levels of the academic self-concept and academic achievement change between grade 3 and grade 5?
3. Are levels of academic self-concept and school grades in grade 3 and grade 5 gender related?
4. Does gender influence mean levels of academic self-concept and academic achievement in grades 3 and 5?
5. Are gender effects related to the longitudinal relationship between academic self-concept and achievement (i.e. autoregressive and cross-lagged coefficients)?

The following hypotheses were set in relation to the research questions:

1. A reciprocal relationship should exist between academic self-concept and academic achievement. It was anticipated that baseline academic self-concept should allow

prediction of achievement at 2-year follow-up (and vice versa). However, it could—in line with the developmental perspective—be that impact from previous academic achievement on later academic self-concept would be the stronger.

2. Both academic self-concept and achievement would tend to attrition.
3. In both waves of the study, girls would demonstrate higher academic self-concept and school grades.
4. With regard to questions 4 and 5, due to the paucity of previous related research and its lack of conclusions, no working hypotheses could be presented for gender influence on (a) change in mean levels of academic self-concept and achievement between grades 3 and 5 and (b) the longitudinal relationship between academic self-concept and academic achievement (i.e. autoregressive and cross-lagged coefficients). In this context, our study was exploratory rather than confirmatory.

Material and methods

Data source and sample

Data was drawn from the larger longitudinal study 'School Effectiveness Research' (Dolata 2014), investigating a representative sample of pupils from Polish elementary schools. The study used a stratified two-stage cluster sampling procedure. The strata were determined by type of urbanisation and the number of class units in a school. Within strata, schools were sampled with a probability proportional to size (number of pupils).

At time 1 (T1), third grade pupils ($N = 4646$; 49.7 % female; ~10-year-olds) completed a general academic self-concept questionnaire and were evaluated by teachers for academic achievement. Two years later (T2), when pupils were in the fifth grade (~12-year-olds), the same children completed the questionnaire and were re-evaluated by teachers. For this study, only participants who had both completed an academic self-concept questionnaire and been evaluated by teachers for academic performance in both grade 3 and 5 were included. The final sample for the current study was 4226 pupils from 270 class units in 157 Polish elementary schools (49.7 % female).

Instruments

Academic self-concept

To measure general academic self-concept (GAS-C), a 15-item Study Motivation (SM) subscale of the *Fragebogen zur Erfassung von Dimensionen der Integration von Schülern* was used (FDI 4–6; Haeberlin et al. 1989). This subscale contains nine positively and six negatively worded items (e.g. 'I am a talented student' or 'Learning is difficult for me'). No item refers to specific school subjects or skills; all require assessment of one's own general academic performance and thus the subscale fits the operational definition of school self-concept. Participants indicated their response on a 4-point scale with anchors of 1 (not true) and 4 (completely true).

The measurement of academic self-concept with only the dimension of school self-concept is aligned with the structure of the curriculum for the first three grades of the Polish school system in which there is no formal division into school subjects. In such a setting, it is

reasonable to assume that at least some children might have difficulties differentiating between more specific aspects of their academic self-concept (Szumski and Karwowski 2015).

In this study, Cronbach's alpha coefficient for the SM subscale of the FDI was 0.868 for grade 3 and 0.910 for grade 5.

Teacher-assigned marks

During the first stage of education, i.e. grades 1–3 in Polish schools, there is no formal division of schooling into subjects and end of term assessment is only descriptive. There are no formal requirements for this description, but in general, they are intended to provide an individual summary for pupils' achievement in all subjects. However, it should not be inferred that pupils at this stage are not given formal grades during the year—such a decision and any rationale for formal grading is at the discretion of the school authority. To obtain data, otherwise not available, teachers were requested to rate each student according to a 4-point scale for Polish language and mathematics.

Teachers classified each pupil as (1) weak—has a weak grasp of the material, makes numerous mistakes and requires systematic assistance; (2) average—deals with requirements but not independent, makes mistakes and needs assistance; (3) performs well, rarely makes mistakes and requires little help; or (4)—excels in all required skills and demonstrates independence in performance of activities.

After third grade at Polish primary schools, each subject is evaluated separately and pupils obtain a final subject grade at the end of each term. Grading is on a criterion-referenced six-degree grade scale: inadequate (1), adequate (2), satisfactory (3), good (4), very good (5) and excellent (6). Grades were obtained from school records at the end of each school year. For comparison of means between the evaluations in grade 3 and grade 5, grades from grade 5 were transformed from six to four levels by combining the scale extrema (1 and 2; 5 and 6).

In further analysis, academic achievement in grade 3 and grade 5 was operationalized as an average grade for language (Polish) and mathematics, with a higher score indicating greater achievement.

Analysis

Preliminary factor analysis of the GAS-C scale

Since factor structure for the GAS-C subscale of the FDI has not yet been established, a series of confirmatory factor analysis (CFA) using the polychoric correlation matrix were used with means and variance adjusted weighted least squares estimators (WLSMV) to account for ordinal responses of GAS-C items.

Three CFA models were tested: unidimensional (M1), two correlated factors (in which all positively worded items defined the first factor and the negatively worded items defined the second factor) (M2) and bifactor models (general academic self-concept factor underlying all variables, as well as two specific dimensions: one for positively worded and the other for negatively worded items; M3). In the bifactor model, all factors were assumed orthogonal, so that domain-specific factors should have captured unique variance of items over and above the general factor, that is, variance not explained by the common factor (Fig. 1).

Model fit was assessed by several commonly used fit indices (Byrne 2011): (1) root mean square error of approximation, RMSEA (Steiger 1990), (2) Tucker-Lewis index, TLI (Tucker
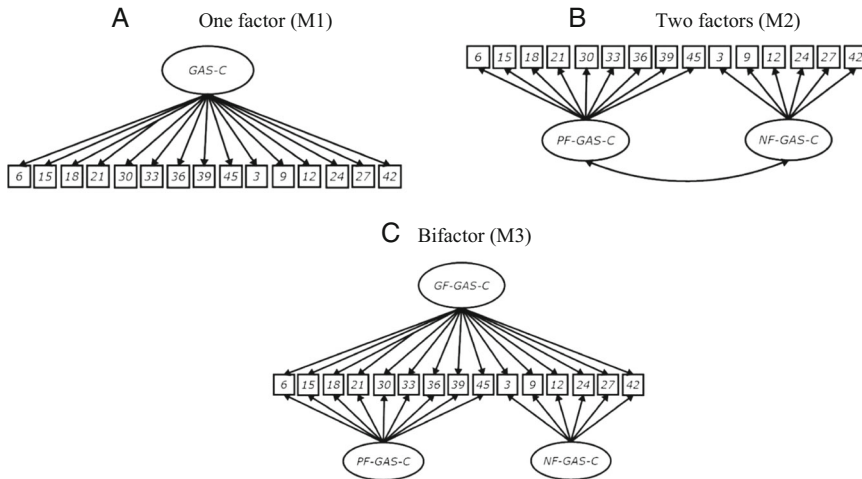
Note. GAS-C = General Academic Self-Concept; GF = General Factor; PF = factor grouping positively worded items; NF = factor grouping negatively worded items.

Fig. 1 **a–c** Graphical representation of models used to test GAS-C scale

and Lewis 1973) and (3) comparative fit index, CFI (Bentler 1990). A model was considered acceptable if RMSEA was less than or equal ro 0.06, and figures for CFI and TLI were close to 0.9 or greater (Marsh et al. 2005a, b; Yu 2002). It was assumed that smaller values of RMSEA and larger values of CFI and TLI would indicate a better fit in comparison of models.

Measurement invariance of the GAS-C scale

Measurement invariance is critical for any longitudinal or multi-group comparison (Millsap 2011). If measurement invariance cannot be established, then a between-time and/or group difference found cannot be interpreted unambiguously (Horn and Mcardle 1992). Following the general procedures proposed by Vandenberg and Lance (2000), this study used a sequential testing procedure (forward approach) to measure invariance. This procedure is based on estimation of a series of hierarchically nested models with increasing constraint of the measurement parameters under test. It starts with the least constrained solution (total lack of invariance) and subsequent restrictions for equality of specific parameters between times and groups are successively imposed. Resulting nested models are then tested against each other and changes to the fit indices analysed (Dimitrov 2010).

The first to be tested was a configural invariance model (M1), i.e. a model with no invariance of parameter estimates for gender and time (i.e. all parameters freely estimated). In the second step, metric invariance (M2) was estimated, restricting equal factor loadings between gender and waves (weak invariance). In the third step, scalar (strong) invariance (M3) was estimated to include restrictions from M2 with the additional constraint of an equal threshold between gender and waves. Residuals for the same items at different study times were correlated. Thus, the longitudinal invariance routine was performed according to the general recommendations of Little (2013).

In order to determine invariance for constrained parameters, following the recommendations of Meade et al. (2008), change for two fit indices between nested models was investigated: CFI and RMSEA. The relatively liberal rule was adopted that the hypothesis of the measurement invariance would be rejected, when the difference (Δ) between the more and less restricted model was lower than −0.002 and for CFI and RMSEA above 0.007 (Meade et al. 2008).

Cross-lagged panel model

In order to examine longitudinal gender relationships between academic self-concept and academic achievement, auto-regressive cross-lagged (ARC-L) modelling with latent variables and multiple indicators were used. The ARC-L model identifies the relationship between variables over time, allowing better understanding of their development. Results from ARC-L models could be used—in tandem with theory—as one element of a larger argument in favour of a causal relationship (Sellig and Little 2012).

In this model, auto-regressive and cross-lagged effects can be analysed (Geiser 2013). The auto-regressive component provides information about relative stability (continuity) of a specified construct over time or—more precisely—about how much variance of the variable X at time T is explained by the same variable measured previously. This implies stability for inter-individual but not necessarily intra-individual differences (Hertzog and Nesselroade 1987). Cross-lagged effects represent the longitudinal prediction of a chosen construct at time T based on another at time $T-1$, controlling for the auto-regressive component.

In the context of the planned analysis, it was particularly important that results from ARC-L could be used to determine whether cross-lagged effects occurred in both directions (i.e. whether X_1 predicted Y_2 and Y_1 predicted X_2) or in only one direction (i.e. X_1 predicted Y_2, but Y_1 did not predict X_2) and to assess the relative strength of cross-lagged effects (i.e. X_1 predicting Y_2 and Y_1 predicting X_2, but impact of X_1 on Y_2 stronger than of Y_1 on X_2). This model allowed direct testing of the hypothesis about the relationship between prior academic self-concept and academic achievement at a later stage.

Software and estimation methods

Analysis was performed using Mplus 7.4 (Muthén and Muthén 1998–2015). Models were estimated using weighted least squares mean and variance adjusted (WLSMV) estimators to account for the ordinal responses of the GAS-C items. Since analysed data were hierarchical—children were nested in classes—the complex sample option in Mplus was used to avoid bias of standard errors and test statistics.

Results

Teacher assessments by gender and grades

Consistent with expectations, girls obtained higher average marks than boys in grade 3 (see Fig. 2). The average for boys was 2.81 (SE = 0.03; $p < 0.001$); for girls, it was higher at 3.02 (SE = 0.03; $p < 0.01$). The difference was statistically significant ($\Delta\text{GPA}_{\text{Girls-Boys}} = 0.21$; SE = 0.03; $p < 0.01$). Results also showed a significant female advantage in grade 5. For girls, M_{GPA} in grade 5 was 2.84 (SE = 0.03) and for boys 2.44 (SE = 0.03; $\Delta\text{GPA}_{\text{Girls-Boys}} = 0.40$,
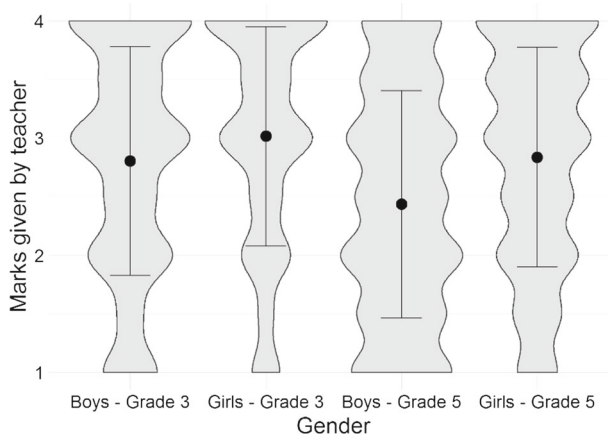
Note. Dots represent mean values, bars represent standard deviation, and shaded area represent an approximate shape of the distribution of marks in groups.

Fig. 2 Violin plot showing mean marks given by teacher across gender and grade

SE $= 0.03$, $p < 0.01$). Although average marks tended to decline over time for both sexes, girls ($\Delta GPA_{Grade5\text{-}Grade3} = -0.18$, SE $= 0.03$, $p < 0.01$) and boys ($\Delta GPA_{Grade5\text{-}Grade3} = -0.37$, SE $= 0.03$, $p < 0.01$), a larger decrease was observed in boys. The mean between gender change in score was significant ($\Delta GPA_{ChangeGrade5\text{-}Grade3} = 0.19$, SE $= 0.02$, $p < 0.01$).

Preliminary factor analysis of the GAS-C scale

The comparison (see Table 1) of the unidimensional, two correlated factors (in which all positively worded items defined the first factor and the negatively worded items defined the second factor) and bifactor models (general academic self-concept factor underlying all variables, as well as two specific dimensions: one for positively worded items and one for negatively worded ones) showed superiority of the bifactor model for both grade 3 and grade 5. These analyses revealed strong support for the 'essential unidimensionality' in Stout's (1987) sense.

Time and gender invariance of general academic self-concept

Model M3 (bifactor) was used as the baseline model for testing multi-group longitudinal measurement invariance (see Table 2). The model with configural invariance showed good model fit (RMSEA $= 0.024$; CFI $= 0.990$; TLI $= 0.988$). The same applied to models with weak (metric) factorial invariance (model 2 in Table 2). When thresholds were constrained (i.e. assessment of strong/scalar invariance), changes in model fit indices were lower than the cutoff value ($\Delta CFI = -0.004$), and the model deemed non-invariant. Examination of the modification indices suggested that model fit would be improved by freeing the parameter for the third threshold for item 45 in grade 3 for girls (model 3A in Table 2), for third (model 3B) and second (model 3C) threshold parameters in grade 5 for boys (also for item 45). After freeing this last threshold, changes in all model fit indices were above the cutoff value ($\Delta RMSEA = 0.002$; $\Delta CFI = -0.002$). By virtue of the

Table 1 Goodness-of-fit of the different models for grade 3 and grade 5

Grade	Models	χ^2	df	P	RMSEA	CFI	TLI
3	One factor (M1)	2839.917	90	<0.001	0.085	0.906	0.890
	Two factors (M2)	1287.450	89	<0.001	0.056	0.959	0.952
	Bifactor (M3)	677.561	75	<0.001	0.044	0.979	0.971
5	One factor (M1)	3375.385	90	<0.001	0.093	0.946	0.936
	Two factors (M2)	1445.789	89	<0.001	0.060	0.978	0.973
	Bifactor (M3)	937.633	75	<0.001	0.052	0.986	0.980

χ^2 chi-square, *df* degrees of freedom, *P* p value, *RMSEA* root mean square error of approximation, *CFI* comparative fit index, *TLI* Tucker-Lewis index

acceptable fit of a model with partially strong/scalar factorial invariance, this model was suitable for use in subsequent analysis.

Longitudinal change for general academic self-concept by gender

Based on the partially invariant bifactor GAS-C model, the latent mean analysis approach estimated the differences between general academic self-concept in boys and girls over time (see Fig. 3). The level of academic self-concept in grade 3 for girls was estimated at 0.02 (SE = 0.03; p = 0.64), compared to 0 for boys (in grade 3). From this, it should be inferred that general academic self-concept in grade 3 did not vary significantly between boys and girls. In grade 5, the average GAS-C for boys was −0.15 (SE = 0.03, p < 0.01) and for girls −0.22 (SE = 0.04, p < 0.01). For both gender groups, self-concept showed a general downward tendency. At the same time, grade 5 girls are characterised by a slightly lower level of general academic self-concept (ΔGAS-C$_{\text{Girls-Boys}}$ = −0.07, SE = 0.03, p = 0.03). This result means that

Table 2 Summary of goodness of fit for models of general academic self-concept—longitudinal measurement invariance for gender

	Model	Par.	test χ^2	df	TLI	CFI	ΔCFI	RMSEA	ΔRMSEA
M1	Configural[a]	336	1576.4	714	0.988	0.990	–	0.024	–
M2	Metric[b]	260	1685.5	790	0.989	0.990	0	0.023	−0.001
M3	Scalar[c]	174	2110.8	876	0.986	0.986	−0.004	0.026	0.003
M3A	Scalar[mod1]	175	2057.3	875	0.987	0.987	−0.003	0.025	0.002
M3B	Scalar[mod2]	176	2026.8	874	0.987	0.987	−0.003	0.025	0.002
M3C	Scalar[mod3]	177	1996.4	873	0.988	0.988	−0.002	0.025	0.002

χ^2 chi-square goodness of fit statistic, *df* degrees of freedom, *TLI* Tucker-Lewis index, *CFI* comparative fit index, *RMSEA* root mean square error of approximation, *mod1* freely third threshold in grade 3 for girls (item 45), *mod2* additionally freed third threshold in grade 5 for boys (item 45), *mod3* additionally freed second threshold in grade 5 for boys (item 45)

[a] Factor loadings and thresholds freely estimated

[b] Factor loadings constrained but thresholds freely estimated

[c] Factor loadings and thresholds constrained
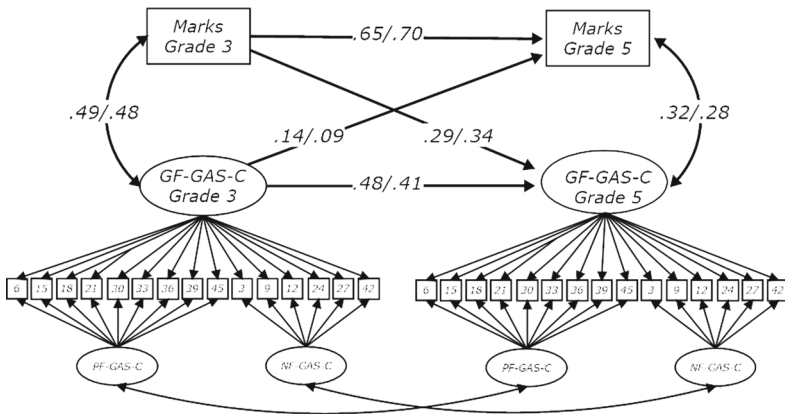
Note. The first number (before slash) is coefficient for boys, the second one (after slash) is coefficient for girls. All coefficients are significant at p <0.01. Correlated uniqueness between corresponding items not been shown for clarity.

Fig. 3 Path diagram (standardised values) for the auto-regressive cross-lagged panel model assuming partially strict measurement invariance for gender and time (see model 3C in Table 2)

although mean GAS-C tended to decline with time for girls and boys, a larger decrease for girls was observed.

Cross-lagged model for academic achievement and general academic self-concept

The multi-group cross-lagged model with autoregressive effects based on the partially scalar invariance bifactor model (see model 3A in Table 2) fitted the data well (RMSEA = 0.026; CFI = 0.984; TLI = 0.984).

Testing for difference parameters to distinguish girls from boys (see last column in Table 3) failed to demonstrate a difference for unstandardised values for auto-regressive, cross-lagged and between-time correlation coefficients (for all $p > 0.05$). Therefore, all the 'structural' parameters in ARC-L model appear similar for either gender.

Analysis indicated that between grade 3 and grade 5, there was a demonstrable, significant impact from earlier academic self-concept on later academic self-concept (girls $\beta = 0.41$; boys $\beta = 0.48$). Also, academic achievement as measured by teachers in school marks assigned in grade 3 significantly influenced marks in grade 5 (for girls $\beta = 0.70$; for boys $\beta = 0.65$). A stronger auto-regressive effect was observed for grades than for self-concept. This means that school marks are more stable over time than academic self-concept (see Table 3 and Fig. 3).

Marks obtained in grade 3 significantly influenced self-concept, as measured in grade 5 (girls $\beta = 0.34$; boys $\beta = 0.29$). The positive value of the coefficient indicates that achievement of better marks resulted in increased self-concept. Also, earlier self-concept showed a relationship with later academic achievement (for girls $\beta = 0.09$; for boys $\beta = 0.14$). The higher the self-concept was in grade 3, the greater the academic achievement 2 years later. However, cross-lagged effects are clearly greater for the impact of marks on academic self-concept than vice versa. It is important to note that all dependencies described so far were not gender dependent.

Table 3 Selected parameters of the ARCL model

Path	Type	Standardised		Unstandardised		
		Boys	Girls	Boys	Girls	ΔGirls–boys
GAS-CT1 → GAS-CT2	AR	0.48 (0.03)**	0.41 (0.03)**	0.48 (0.03)**	0.39 (0.04)**	−0.09 (0.05) ns
TMT1 → TMT2	AR	0.65 (0.02)**	0.70 (0.02)**	0.65 (0.02)**	0.70 (0.03)**	0.05 (0.04) ns
GAS-CT1 → TMT2	CL	0.14 (0.03)**	0.09 (0.03)**	0.14 (0.03)**	0.09 (0.03)**	−0.05 (0.04) ns
TMT1 → GAS-CT2	CL	0.29 (0.03)**	0.34 (0.03)**	0.27 (0.03)**	0.32 (0.03)**	0.05 (0.04) ns
GAS-CT1 ↔ TMT1	Cor	0.49 (0.03)**	0.48 (0.03)**	0.45 (0.04)**	0.41 (0.03)**	−0.04 (0.05) ns
GAS-CT2 ↔ TMT2	Cor	0.32 (0.02)**	0.28 (0.02)**	0.14 (0.01)**	0.11 (0.01)**	−0.03 (0.02) ns
GAS-CT1 (M)	M	–	–	0Fix	0.03 (0.05) ns	–
GAS-CT2 (I)	I	–	–	−0.84 (0.09)**	1.12 (0.09)**	−0.28 (0.10)**
TMT1 (M)	M	–	–	2.81 (0.02)**	3.02 (0.03)**	0.21 (0.04)**
TMT2 (I)	I	–	–	0.60 (0.07)**	0.73 (0.07) **	0.13 (0.10) ns

GAS-C General Academic Self-Concept, *TM* teacher marks, *T1* grade 3, *T2* grade 5, *AR* auto-regressive, *CL* cross-lagged, *Cor* correlation, *M* mean, *I* intercept, *Fix* fixed at 0 for model identification, Δ difference, → regression, ↔ correlation, *ns* non-significant

$*p < 0.05$; $**p < 0.01$

Discussion

Interpretation of results

The first objective of this study was to investigate the causal longitudinal relationship hypothesised between general academic self-concept and academic achievement. Consistent with prior studies and expectations (hypothesis 1), both constructs—after controlling for significant auto-lagged effects—were found to be reciprocally influenced over time. Further, the results extended prior research by demonstrating that cross-lagged paths from prior academic achievement to subsequent self-concept were stronger than the paths from academic self-concept to academic achievement (e.g. Chapman and Tunmer 1997; Hoge et al. 1995; Muijs 1997).

The relationship between prior achievement and subsequent academic self-concept requires comment in the context of the grading system at the early stages of education in Poland. In general, two mechanisms are offered to explain the influence of school grades on academic self-concept. One is that grades form the basis for pupils' social comparisons. The second treats grades as a source for feedback with regards to ability (Marsh 1984). However, teacher ratings in grade 3 were not available to students and, as such, could not, directly, fit either role required by these explanations. To demonstrate any influence on academic self-concept, the ratings must correlate with something assuming that role. Some hypotheses are proposed here for this relationship.

It is clear that teachers based their ratings on an image projected by overall academic performance. This image might be construed with the help of the formal grade system employed by teachers during the year. In this case, the grade system might perform such a role, as demanded by the mechanisms, previously mentioned. Since some type of formal grade

system is applied in most schools, this might suffice as explanation for the observed relationships. An alternative view, however, can be postulated.

In line with the Shavelson model of self-concept, grades might form just one source of feedback for students (Shavelson et al. 1976). Research shows that instructions or feedback, similar to marks, influence pupil self-concept and thereby motivation to study and form part of the learning process (cf. Lazarides and Ittel 2012; Wentzel et al. 2010).

It is then possible that academic self-concept is dynamically reinforced through experience of daily school life. In this case, feedback during lessons, as well as the dynamics of student–teacher interaction, might play some formative role for academic self-concept. The observed relationship between teacher ratings (and, more generally, grades) would then complement the mechanism described. The extent to which ratings and grades are effective in this role would influence their validity as predictors of academic self-concept. Unfortunately, the data analysed here does not allow verification of this hypothesis.

The results of the present study also support the hypothesis that both academic self-concept and achievement decline with time (hypothesis 2). Attrition of academic self-concept confirmed the results of studies indicating that children in the first grades of primary school are characteristically overoptimistic in terms of academic self-concept (Wigfield et al. 1997). This is related to their limited ability to discriminate between qualities they actually possess and those considered desirable (Lindberg et al. 2013), in addition to ego-centrism (Marsh 1985), which hinders capacity to exploit social comparison mechanisms for accurate self-evaluation (Anderman and Maehr 1994; Harter 1999).

During academic development, academic self-concept becomes more accurate (Wigfield et al. 2006). It is increasingly performance-oriented and based on frequent competence-related feedback. However, at the same time children become increasingly aware of their abilities and can also evaluate quality of their performance without an external feedback (Dweck 2002). At the age of 10–12, many children start to identify their skills with grades obtained and identify failure with a low level of skill (for research review, see Dweck 2002).

Only limited support, however, was found for the influence of gender hypothesised for changes to mean levels of academic self-concept or achievement between grades 3 and 5 (hypothesis 3). It was predicted that at both ages, girls would demonstrate higher academic self-concept and school grades than boys. In line with the assumptions, girls characteristically obtained higher mean grades in both grades 3 and 5. Unexpectedly, though, we found that the mean level of girls' academic self-concept did not vary from boys' in grade 3 and it decreased more rapidly with time. This finding is even more surprising, since no gender effects related to the longitudinal relationship between academic self-concept and academic achievement were observed (i.e. autoregressive or cross-lagged coefficients). This points to the active presence of gender-related determinants for academic self-concept.

It is possible that, for boys, academic self-concept is more strongly related with marks for subjects outside those analysed and that this relationship performed a compensatory role. In particular, it might be expected that there are school subjects in which boys are rewarded with higher grades than girls and which preferentially influence their academic self-concept (e.g. sport). Simultaneously, the extent of decrease in academic self-concept in girls may be influenced by factors similar to those responsible for a greater decrease in general female self-concept at this age (e.g. changes related to puberty; Robins and Trzasniewski 2005). Alternatively, marks obtained by boys and girls might be accompanied by differentiated feedback from teachers. In order to observe such an effect, boys would have to receive information that boosts their academic self-concept relative to girls.

Research implications

In different countries, formal evaluation of students (i.e. grading systems) is introduced at various stages. Assuming that grades awarded by teachers shape academic self-concept, it would be relevant to check whether earlier introduction of a grading system encourages more rapid self-concept crystallisation.

If confirmed, this hypothesis would explain why results of certain studies in countries where introduction of school grading is 'delayed' resonate more with the skill development model than the reciprocal effects or the self-enhancement models (e.g. Skaalvik 1997; Skaalvik and Valås 1999). In this context, the stronger impact of previous school grades on subsequent self-concept rather than vice versa, as found in our study, might be explained as symptomatic of education system structure.

Further, collection of observational data is recommended to explain the mechanism of translation of grades to self-concept as well as to address the lack of gender differences with respect to average levels of self-concept. Observation should focus on detailed logging of instructions and feedback from teachers.

Practical implications

Impact of grading on student self-concept, in spite of being indirect, indicates that pupils evaluate their own skills based on other information transmitted by teachers during classes. Therefore, for reason of practical implications, the significance of feedback in shaping school children's academic self-concept should be strongly emphasised.

The primary task for teacher should be to supply feedback to pupils in a manner that does not damage self-concept. Strengthening self-concept potentially improves motivation to learn and involvement with the learning process. As emphasised by Hattie and Timperley (2007), efficient feedback implies providing students with specific information about what they do right and wrong and what requires improvement or is achieved differently than previously. It is necessary to avoid negative social comparison of accomplishment and to emphasise achievement on a time scale as well as to reinforce positive peer relations and pro-social behaviour in class (Manning 2007).

Limitations

The current study was subject to certain limitations:

Firstly, grades were the only measure of academic achievement referenced by the study. The present research did not investigate how more objective measures of academic achievement, such as standardised test scores, might influence academic self-concept. Secondly, academic achievement in grades 3 and 5 was only measured for pupils' native Polish language and mathematics, by grade average. In future studies, it would be desirable to increase the number of domains evaluated for academic achievement.

Thirdly, the study measured general academic self-concept rather than subject-specific academic self-concept. Future work might focus on academic self-concept for specific school subjects.

Fourthly, results relied only on two study waves. Future research should include data from more stages of education. Consideration of more than two waves of study would allow more complex statistical methods (Reinecke and Seddig 2011), e.g. the latent growth curve model (Duncan and Duncan 2004; Bollen and Curran 2006). Finally, our study did not consider

potential moderators or mediators in the relation between academic self-concept and academic achievement, e.g. motivation to learn, self-efficacy, locus of control, attribution of successes and failures, implicit intelligence theories (cf. Wigfield et al. 2006) or level of optimism (Wesson and Derrer-Rendall 2011).

Despite these limitations, we anticipate that this study will shed more light on the longitudinal relationship between general academic self-concept and academic achievement with particular regard to gender-specific effects.

Final remarks

Analysis revealed that between grades 3 and 5 in Polish primary schools, pupil attainment, as measured by marks awarded by teachers, and academic self-concept both declined. This effect was observed for both boys and girls. Further, academic achievement and academic self-concept were reciprocally influenced over time, but influence from prior achievement on self-concept was stronger.

Such a set of relations between the two constructs suggests that delaying formal assessment, in the form of teacher-assigned marks, might lead to slower crystallisation of academic self-concept. Whether such effect is more beneficial in the long or short term—either for academic achievement or personal development—remains a separate issue.

The present study suggests that even if pupils are unaware of their scores which, therefore, do not permit social comparison, assessment does influence academic self-concept. This implies that information about pupil performance is transmitted by the teacher through alternative channels, e.g. feedback. This highlights the importance of the pupil–teacher interaction in the context of academic self-concept research. From the teacher perspective, such interaction might establish the approach adopted for assessment of skills, which then translates into a mark. From the pupil view, feedback is both a foundation and an important factor for development of academic self-concept.

This study highlights the need for special focus on the role and quality of teacher feedback, and its importance as an issue which should be covered during teacher training. This should respect feedback's potentially diverse influence on academic self-concept for boys and girls. The ultimate goal should be pupil–teacher interaction to enhance each pupil's academic self-concept and perseverance, providing motivation to embrace active learning and greater educational achievement.

References

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research, 64*(2), 287–309. doi:10.3102/00346543064002287.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. doi:10.1037/0033-2909.107.2.238.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: a structural equation perspective*. Hoboken, N.J: Wiley-Interscience.

Burusic, J., Babarovic, T., & Seric, M. (2012). Differences in elementary school achievement between girls and boys: does the teacher gender play a role? *European Journal of Psychology of Education, 27*(4), 523–538. doi:10.1007/s10212-011-0093-2.

Byrne, B. M. (1984). The general/academic self-concept nomological network: a review of construct validation research. *Review of Educational Research, 54*(3), 427–456. doi:10.3102/00346543054003427.

Byrne, B. M. (1986). Self-concept/academic achievement relations: an investigation of dimensionality, stability, and causality. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement, 18*(2), 173–186. doi:10.1037/h0079982.

Byrne, B. M. (2011). *Structural equation modeling with Mplus: basic concepts, applications, and programming* (1st ed.). New York: Routledge.

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: cause or effect of academic achievement? *Journal of Educational Psychology, 69*(2), 136–145. doi:10.1037/0022-0663.69.2.136.

Chang, L., McBride-Chang, C., Stewart, S., & Au, E. (2003). Life satisfaction, self-concept, and family relations in Chinese adolescents and children. *International Journal of Behavioral Development, 27*(2), 182–189. doi:10.1080/01650250244000182.

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology, 67*(3), 279–291. doi:10.1111/j.2044-8279.1997.tb01244.x.

Conger, D., & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The ANNALS of the American Academy of Political and Social Science, 627*(1), 184–214. doi:10.1177/0002716209348751.

Craven, R. G., & Marsh, H. W. (2008). The centrality of the self-concept construct for psychological wellbeing and unlocking human potential: implications for child and educational psychologists. *Educational and Child Psychology, 25*(2), 104–118.

Craven, R. G., McInerney, V., & Marsh, H. W. (2000). The structure and development of young children's self-concepts and relation to academic achievement (pp. 1–20). Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149. doi:10.1177/0748175610373459.

Dolata, R. (Ed.). (2014). *Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym* (Vol. 1–2). Warszawa: Instytut Badań Edukacyjnych.

Downey, D. B., & Vogt Yuan, A. S. (2005). Sex differences in school performance during high school: puzzling patterns and possible explanations. *The Sociological Quarterly, 46*(2), 299–321. doi:10.1111/j.1533-8525.2005.00014.x.

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198–208. doi:10.1037/0022-0663.98.1.198.

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy, 35*(2), 333–363. doi:10.1016/S0005-7894(04)80042-X.

Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 57–88). San Diego: Academic. **Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B978012750053950005X**.

Eccles, J. S. (1993). School and family effects on the ontogeny of children's interests, self-perceptions, and activity choices. In J. E. Jacobs (Ed.), *Developmental perspectives on motivation* (Vol. 40, pp. 145–208). Lincoln and London: University of Nebraska Press.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task self-perceptions during elementary school. *Child Development, 64*(3), 830–847. doi:10.1111/j.1467-8624.1993.tb02946.x.

Ekstrom, R. B. (1994). Gender differences in high school grades: an exploratory study. *College Board Report, 94*(3), 1–30.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin, 136*(1), 103–127. doi:10.1037/a0018053.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research, 59*(4), 395–430. doi:10.3102/00346543059004395.

Fischer, F. T., Schult, J., & Hell, B. (2013). Sex-specific differential prediction of college admission tests: a meta-analysis. *Journal of Educational Psychology, 105*(2), 478–488. doi:10.1037/a0031956.

Fraine, B. D., Damme, J. V., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: a multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*(1), 132–150. doi:10.1016/j.cedpsych.2006.10.005.

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*(4), 519–533. doi:10.1037/0012-1649.38.4.519.

Geiser, C. (2013). *Data analysis with Mplus*. New York: The Guilford Press.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*(1), 124–136. doi:10.1037/0022-0663.95.1.124.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: a ten-year longitudinal study. *Self and Identity, 3*(1), 53–68. doi:10.1080/13576500342000040.

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: mediating and additive effects. *Learning and Individual Differences, 20*(6), 644–653. doi:10.1016/j.lindif.2010.08.001.

Hadjar, A., Krolak-Schwerdt, S., Priem, K., & Glock, S. (2014). Gender and educational achievement. *Educational Research, 56*(2), 117–125. doi:10.1080/00131881.2014.898908.

Haeberlin, U., Moser, U., Bless, G., & Klaghofer, R. (1989). *Integration in die schulklasse: fragebogen zur erfassung von dimensionen der integration von schülern: FDI 4–6*. Bern, Stuttgart: Verlag Paul Haupt.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*(1), 123–142. doi:10.3102/00346543052001123.

Harter, S. (1999). *The construction of the self: a developmental perspective*. New York: Guilford Press.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi:10.3102/003465430298487.

Helmke, A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: a longitudinal study. *Journal of Educational Psychology, 87*(4), 624–637. doi:10.1037/0022-0663.87.4.624.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology, 26*(3), 276–295. doi:10.1016/j.appdev.2005.02.007.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: some implications of the trait-state distinction for the structural modeling of developmental change. *Child Development, 58*(1), 93–109. doi:10.2307/1130294.

Hoge, D. R., Smit, E. K., & Crist, J. T. (1995). Reciprocal effects of self-concept and academic achievement in sixth and seventh grade. *Journal of Youth and Adolescence, 24*(3), 295–314. doi:10.1007/BF01537598.

Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117–144. doi:10.1080/03610739208253916.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin, 107*(2), 139–155. doi:10.1037/0033-2909.107.2.139.

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*(2), 509–527. doi:10.1111/1467-8624.00421.

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11–21. doi:10.1016/j.lindif.2013.12.003.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*(2), 198–214. doi:10.1037/0033-2909.105.2.198.

Kling, K. C., Noftle, E. E., & Robins, R. W. (2013). Why do standardized tests underpredict women's academic performance? The role of conscientiousness. *Social Psychological and Personality Science, 4*(5), 600–606. doi:10.1177/1948550612469038.

Lazarides, R., & Ittel, A. (2012). Instructional quality and attitudes toward mathematics: do self-concept and interest differ across students' patterns of perceived instructional quality in mathematics classrooms? *Child Development Research, 2012*, 11. doi:10.1155/2012/813920.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135. doi:10.1037/a0021276.

Lindberg, S., Linkersdörfer, J., Ehm, J.-H., Hasselhorn, M., & Lonnemann, J. (2013). Gender differences in children's math self-concept in the first years of elementary school. *Journal of Education and Learning, 2*(3). http://doi.org/10.5539/jel.v2n3p1.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: The Guilford Press.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*(4), 357–365. doi:10.1177/0165025407077757.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280–296. doi:10.1037/0021-9010.67.3.280.

Manning, M. A. (2007). Self-concept and self-esteem in adolescents. *Student Services*, 11–15.

Marsh, H. W. (1984). Self-concept, social comparison, and ability grouping: a reply to Kulik and Kulik. *American Educational Research Journal, 21*(4), 799–806. doi:10.3102/00028312021004799.

Marsh, H. W. (1985). Age and sex effects in multiple dimensions of preadolescent self-concept: a replication and extension. *Australian Journal of Psychology, 37*(2), 197–204. doi:10.1080/00049538508256398.

Marsh, H. W. (1986). Verbal and math self-concepts: an internal/external frame of reference model. *American Educational Research Journal, 23*(1), 129–149. doi:10.3102/00028312023001129.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: preadolescence to early adulthood. *Journal of Educational Psychology, 81*(3), 417–430. doi:10.1037/0022-0663.81.3.417.

Marsh, H. W. (1990a). A multidimensional, hierarchical model of self-concept: theoretical and empirical justification. *Educational Psychology Review, 2*(2), 77–172. doi:10.1007/BF01322177.

Marsh, H. W. (1990b). Causal ordering of academic self-concept and academic achievement: a multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*(4), 646–656. doi:10.1037/0022-0663.82.4.646.

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal, 28*(2), 445–480. doi:10.3102/00028312028002445.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: the role of self concept in educational psychology.* Leicester: British Psychological Society.

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology, 95*(4), 687–706. doi:10.1037/0022-0663.95.4.687.

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*(2), 133–163. doi:10.1111/j.1745-6916.2006.00010.x.

Marsh, H. W., & Gouvernet, P. J. (1989). Multidimensional self-concepts and perceptions of control: construct validation of responses by children. *Journal of Educational Psychology, 81*(1), 57–69. doi:10.1037/0022-0663.81.1.57.

Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: reunification of east and west German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology, 29*(3), 264–282. doi:10.1016/S0361-476X(03)00034-1.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: relations and causal ordering. *British Journal of Educational Psychology, 81*(1), 59–77. doi:10.1348/000709910X503501.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: its multifaceted, hierarchical structure. *Educational Psychologist, 20*(3), 107–123. doi:10.1207/s15326985ep2003_1.

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: relations to academic self-concept and achievement. *American Educational Research Journal, 34*(4), 691–720. doi:10.3102/00028312034004691.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-description questionnaire: age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology, 76*(5), 940–956. doi:10.1037/0022-0663.76.5.940.

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*(3), 366–380. doi:10.1037/0022-0663.80.3.366.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: measurement and multidimensional structure. *Journal of Educational Psychology, 83*(3), 377–392. doi:10.1037/0022-0663.83.3.377.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181–220. doi:10.1207/s15327906mbr3302_1.

Marsh, H. W., Hau, K.-T., & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement: influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39*(3), 727–763. doi:10.3102/00028312039003727.

Marsh, H. W., Parada, R. H., & Ayotte, V. (2004). A multidimensional perspective of relations between self-concept (Self Description Questionnaire II) and adolescent mental health (Youth Self-Report). *Psychological Assessment, 16*(1), 27–41. doi:10.1037/1040-3590.16.1.27.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005a). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares, J. J. McArdle, & R. P. McDonald (Eds.), *Contemporary psychometrics: a festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah: Erlbaum.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005b). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development, 76*(2), 397–416. doi:10.1111/j.1467-8624.2005.00853.x.

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*(4), 311–360. doi:10.1207/s15327574ijt0604_1.

Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, O., & Brettschneider, W.-D. (2007). Longitudinal study of preadolescent sport self-concept and performance: reciprocal effects and causal ordering. *Child Development, 78*(6), 1640–1656. doi:10.1111/j.1467-8624.2007.01094.x.

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: a synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, vol 1: theories, constructs, and critical issues* (pp. 427–458). Washington: American Psychological Association.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: age-cohort and cross-cultural differences. *American Educational Research Journal, 52*(1), 168–202. doi:10.3102/0002831214549453.

Mead, S. (2006). The evidence suggests otherwise: the truth about boys and girls. Retrieved from http://www.ldonline.org/article/19236/.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. doi:10.1037/0021-9010.93.3.568.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Psychology Press.

Muijs, R. D. (1997). Predictors of academic achievement and academic self-concept: a longitudinal perspective. *British Journal of Educational Psychology, 67*(3), 263–277. doi:10.1111/j.2044-8279.1997.tb01243.x.

Mullis, I. V. S., Martin, M., Foy, P., & Drucker, K. (2012). *PIRLS 2011 International Results in Reading*. TIMSS & PIRLS International Study Center & International Association for the Evaluation of Educational Achievement (IEA).

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Järvinen, L. (2012). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability: gender differences and temperament. *British Journal of Educational Psychology, 82*(2), 185–206. doi:10.1111/j.2044-8279.2010.02017.x.

Muthén, B. O., & Muthén, L. K. (1998–2015). *Mplus user's guide. Seventh edition.* 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., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: different countries, different trajectories? *Journal of Research on Adolescence, 20*(2), 482–506. doi:10.1111/j.1532-7795.2010.00644.x.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: a longitudinal study. *Journal of Educational Psychology, 76*(5), 857–873. doi:10.1037/0022-0663.76.5.857.

O'Mara, A. J., & Marsh, H. W. (2006). Support for a reciprocal effects model of self-concept and academic achievement through a contrast of multidimensional and unidimensional approaches. Presented at the Fourth International Biennial SELF Research Conference, Ann Arbor, US.

O'Mara, A. J., Green, J., & Marsh, H. W. (2006a). Administering self-concept interventions in schools: no training necessary? A meta-analysis. *International Education Journal, 7*(4), 524–533.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006b). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*(3), 181–206. doi:10.1207/s15326985ep4103_4.

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress three decades of student performance in reading and mathematics*. Washington: National Center for Education Statistics.

Quirk, M., Schwanenflugel, P., & Webb, M. (2009). A short-term longitudinal study of the relationship between motivation to read and reading fluency skill in second grade. *Journal of Literacy Research, 41*(2), 196–227. doi:10.1080/10862960902908467.

Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *AStA Advances in Statistical Analysis, 95*(4), 415–434. doi:10.1007/s10182-011-0171-4.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. doi:10.1037/a0026838.

Robins, R. W., & Trzesniewski, K. H. (2005). Self-esteem development across the lifespan. *Current Directions in Psychological Science, 14*(3), 158–162.

Rosen, J. A. (2010). Academic Self-Concept. In J. A. Rosen, E. J. Glennie, D. B. W., J. M. Lennon, & R. N. Bozick (Eds.), *Noncognitive skills in the classroom : new perspectives on education research* (pp. 117–144). Research Triangle Park, NC: RTI International.

Sellig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. P. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York: Guilford Press.

Shapka, J. D. (2009). Trajectories of math achievement and perceived math competence over high school and postsecondary education: effects of an all-girl curriculum in high school. *Educational Research and Evaluation, 15*(6), 527–541. doi:10.1080/13803610903354775.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. doi:10.3102/00346543046003407.

Shibley Hyde, J., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly, 25*(4), 364–378. doi:10.1111/1471-6402.00035.

Skaalvik, E. M. (1997). Issues in research on self-concept. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 51–98). Greenwich: JAIPress.

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: a longitudinal study. *The Journal of Experimental Education, 67*(2), 135–149. doi:10.1080/00220979909598349.

Spilt, J. L., Koomen, H. M. Y., & Jak, S. (2012). Are boys better off with male and girls with female teachers? A multilevel investigation of measurement invariance and gender match in teacher–student relationship quality. *Journal of School Psychology, 50*(3), 363–378. doi:10.1016/j.jsp.2011.12.002.

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180. doi:10.1207/s15327906mbr2502_4.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589–617. doi:10.1007/BF02294821.

Szumski, G., & Karwowski, M. (2015). Emotional and social integration and the big-fish-little-pond effect among students with and without disabilities. *Learning and Individual Differences, 43*, 63–74. doi:10.1016/j.lindif.2015.08.037.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. doi:10.1007/BF02291170.

Vail, K. (2006). Is the boy crisis real? *American School Board Journal, 193*(10), 22–23.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: a meta-analytic review. *Educational Psychologist, 39*(2), 111–133. doi:10.1207/s15326985ep3902_3.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. doi:10.1177/109442810031002.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. doi:10.1037/a0036620.

Wang, J. (2006). An empirical study of gender difference in the relationship between self-concept and mathematics achievement in a cross-cultural context. *Educational Psychology, 26*(5), 689–706. doi:10.1080/01443410500390863.

Wang, M.-T., & Eccles, J. S. (2012). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence, 22*(1), 31–39. doi:10.1111/j.1532-7795.2011.00753.x.

Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology, 35*(3), 193–202. doi:10.1016/j.cedpsych.2010.03.002.

Wesson, C. J., & Derrer-Rendall, N. M. (2011). Self-beliefs and student goal achievement. *Psychology Teaching Review, 17*(1), 3–12.

Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement situations. *Educational Psychologist, 26*(3–4), 233–261. doi:10.1207/s15326985ep2603&4_3.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: a 3-year study. *Journal of Educational Psychology, 89*(3), 451–469. doi:10.1037/0022-0663.89.3.451.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In W. Damon & R. M. Lerner (Eds.), *Social, emotional, and personality development* (6th ed., Vol. 3, pp. 933–1002). Hoboken: John Wiley.

Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles: University of California.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111–125. doi:10.1177/1745691613518075.

Author details

Paweł Grygiel (Educational Research Institute, ul. Górczewska 8, 01–180 Warsaw, Poland). Email: p.grygiel@ibe.edu.pl, http://www.ibe.edu.pl/en/

Current themes of research:

Psychometrics, peer relationships, social development, school effectiveness research.

Most relevant publications:

1. Grygiel, P., Humenny, G., Rebisz, S., Świtaj, P., & Sikorska, J. (2013). Validating the Polish Adaptation of the 11-Item De Jong Gierveld Loneliness Scale. *European Journal of Psychological Assessment, 29*(2), 129–139.

2. Grygiel, P., Humenny, G., Rębisz, S., Bajcar, E., & Świtaj, P. (2014). Peer Rejection and Perceived Quality of Relations With Schoolmates Among Children With ADHD. *Journal of Attention Disorders*. http://doi.org/10.1177/1087054714563791

3. Grygiel, P., & Humenny, G. (2015). Efektywność nauczania a relacje rówieśnicze. W R. Dolata, A. Hawrot, G. Humenny, A. Jasińska-Maciążek, M. Koniewski, P. Majkut, … A. Otręba-Szklarczyk (Red.), *(Ko)warianty efektywności nauczania. Wyniki badania w klasach IV-VI* (s. 85–116). Warszawa: Instytut Badań Edukacyjnych.

4. Dolata, R., Grygiel, P., Jankowska, D. M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., … Pisarek, J. (2015). *Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV – VI*. Warszawa: Instytut Badań Edukacyjnych.

5. Grygiel, P. (2015). Test ukrytej struktury kwestionariusza „Poczucie Integracji Rówieśniczej". Częściowo konfirmacyjny model podwójnego czynnika z ładunkami krzyżowymi. W B. Niemierko & M. K. Szmigel (Red.), *Zastosowania diagnozy edukacyjnej* (s. 438–458). Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.

Michał Modzelewski (Educational Research Institute, ul. Górczewska 8, 01–180 Warsaw, Poland). Email: m.modzelewski@ibe.edu.pl, http://www.ibe.edu.pl/en/

Current themes of research:

Psychometrics and school effectiveness research, with special focus on class size effect.

Most relevant publications:

1. Jasińka-Maciążek, A., Modzelewski, M.(2014). Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. Edukacja, 2, 85–107.

2. Jasińka-Maciążek, A., Modzelewski, M. (2014). Pomiar wyników nauczania w zakresie czytania, świadomości językowej i matematyki. W R. Dolata(Red.), Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym (s.27-62). Warszawa: Instytut Badań Edukacyjnych.

3. Jasińka-Maciążek, A., Modzelewski, M. (2014). Międzyszkolne i międzyoddziałowe zróżnicowanie wyników nauczania. W R. Dolata(Red.), Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym (s.93-109). Warszawa: Instytut Badań Edukacyjnych.

4. Modzelewski, M. (2015). Co umieją szóstoklasiści i jak to zmierzono? W R. Dolata, Grygiel, P., Jankowska, D.M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., Modzelewski, M., Pisarek, J. (2015). Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV – VI (s.10-23). Warszawa: Instytut Badań Edukacyjnych.

5. Modzelewski, M. (2015). Czy wielkość klasy ma znaczenie? W R. Dolata, Grygiel, P., Jankowska, D.M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., Modzelewski, M., Pisarek, J. (2015). Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV – VI (s.49-58). Warszawa: Instytut Badań Edukacyjnych.

Jolanta Pisarek (Educational Research Institute, ul. Górczewska 8, 01–180 Warsaw, Poland). Email: j.pisarek@ibe.edu.pl, http://www.ibe.edu.pl/en/

Current themes of research:

School effectiveness, class climate, effective feedback.

1. Hawrot, A., & Pisarek, J. (2014). Inteligencja a wyniki nauczania. W R. Dolata(Red.), Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym (s.111-130). Warszawa: Instytut Badań Edukacyjnych.

2. Pisarek, J., Grygiel, P. (2015). Czy warto dbać o samoocenę szkolną uczniów? W R. Dolata, Grygiel, P., Jankowska, D.M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., Modzelewski, M., Pisarek, J. (2015). Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV – VI (s.72-82). Warszawa: Instytut Badań Edukacyjnych.

3. Pisarek, J., Jarnutowska, E., Grygiel, P. (2015). Jak udzielać informacji zwrotnej? W R. Dolata, Grygiel, P., Jankowska, D.M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., Modzelewski, M., Pisarek, J. (2015). Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV – VI (s.85-94). Warszawa: Instytut Badań Edukacyjnych.