

Gestaltung des Titelblattes der Dissertation
 (§ 8 Promotionsordnung der Fakultät Rehabilitationswissenschaften)
 Anl. 2 PromO

Validation and Evaluation of the Strengths and Difficulties Questionnaire in the German

 School System: An Examination of Regular and Inclusive Classrooms

 (Titel)

vorgelegt von
 Jeffrey Matthew DeVries

 (Name)

als Dissertation zur Erlangung des Grades eines Doktors
 der Philosophie (Dr. phil.)
 in der
 Fakultät Rehabilitationswissenschaften
 der Technischen Universität Dortmund

Dortmund

 2019

 (Jahr)

Betreuer(in): Prof. Dr. Markus Gebhardt.....

Betreuer(in): Prof Dr. Jörg-Tobias Kuhn

This dissertation was submitted to and accepted by the faculty of rehabilitation sciences at the Technische Universität Dortmund.

Referees:

Prof Dr. Markus Gebhardt

Prof. Dr. Jörg-Tobias Kuhn

Defended on 06.03.2020

Table of Contents

1.0 Introduction and Overview	1
2.0 Theoretical Background.....	2
2.1 Internalizing and Externalizing Behaviors	2
2.2 The Strengths and Difficulties Questionnaire	3
2.2.1 Classroom use of the SDQ.....	4
2.2.2 Issues with the SDQ	6
2.3 Social Behavior.....	6
2.4 Inclusion in the Classroom.....	7
2.5 Perception of Inclusion Questionnaire	8
2.6 Research Questions	9
2.7 Overview of the Articles Included in This Framework	10
3.0 Article One: “An assessment of measurement invariance in the 3- and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study	12
3.1 Purpose and Motivation.....	12
3.2 Study Description.....	12
3.3 Results and Discussion	13
3.4 Author Contributions	14
4.0 Article Two: “How Does Social Behavior Relate to Both Grades and Achievement Scores?”	15
4.1 Purpose and Motivation.....	15
4.2 Study Description.....	15
4.3 Results and Discussion	16
4.4 Author Contributions	17
5.0 Article Three: “Do learners with special education needs really feel included? Evidence from the Perception of Inclusion Questionnaire and Strengths and Difficulties Questionnaire”	19
5.1 Purpose and Motivation.....	19

5.2 Study Description.....	19
5.3 Results and Discussion	20
5.4 Author Contributions	21
6.0 General Discussion.....	22
6.1 Review of general results	22
6.1.1 Validation of the SDQ.....	22
6.1.2 Longitudinal change in SDQ scores	23
6.1.3 Relationship of the SDQ to Academic Achievement	23
6.1.4 Inclusion and the SDQ.....	25
6.2 Limitations & Future Work.....	26
6.3 Conclusion.....	26
7.0 References.....	28

List of Tables and Figures

Table 2.1 The Strengths and Difficulties Questionnaire	5
Table 2.2 The Perception of Inclusion Questionnaire	9
Figure 4.1 Path Model for Social Behavior and Achievement	18

1.0 Introduction and Overview

Throughout late childhood and early adolescence, different patterns of internalizing and externalizing behaviors affect development. While these behaviors are often low-intensity or not problematic for many children and adolescents, they can become problematic and relate to academic issues in others. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; 2001) is a popular instrument for identifying the frequency of such problematic behaviors. Herein, I examine the structure and invariance of the SDQ, its efficacy for longitudinal comparisons, its relationship to academic outcomes, and its relationship to inclusion within the classroom.

In the following sections, I explore the theoretical background of the SDQ and its theoretical structure and performance in various regions and settings. Particular care is taken to examine the functioning of the instrument in a school setting, where the relationship between social behavior based on the SDQ's social scales to academic achievement is explored. Next, the relationship of the SDQ to self-reported perceptions of inclusion is discussed. This is followed by descriptions of the three articles which form the core of this work. After discussing each article individually, overall findings and conclusions about the SDQ and inclusion are discussed.

2.0 Theoretical Background

2.1 Internalizing and Externalizing Behaviors

Achenbach's (1966; Achenbach & Edelbrock, 1978; Achenbach, Ivanova, Rescorla, Turner, & Althoff, 2016) classification of behaviors described internalizing and externalizing as two different categories of behavioral issues. He described problematic externalizing behaviors as outward behaviors and internalizing as inward. For example, lashing out, fighting, yelling, physical agitation, and similar behaviors represent externalizing, while expressions of sadness, social withdrawal, lack of motor responses, and the like represent internalizing. Achenbach et al. (2016) and many others (Beauchaine & Crowell, 2018; Hostinar & Cicchetti, 2018) have related these behaviors to current and future disorders. Although these are often thought of as separate dimensions of behavior, some researchers have noted that both can coincide within the same individual. In other words, they are not mutually exclusive behaviors (Oland & Shaw, 2005).

Internalizing and externalizing behaviors are related to the course of childhood development. Although they are often connected to developmental disorders (e.g., Beauchaine & Crowell, 2018; Hostinar & Cicchetti, 2018; Soto-Sanz et al., 2019; Theunissen, Wolff, & Reijneveld, 2019) there is a typical path of development for these behaviors. Many past studies have identified a marked increase in these behaviors during prepubescence and through adolescence (Becker, Rothenberger, Sohn, Ravens-Sieberer, & Klasen, 2015; van der Ende, Verhulst, &

Tiemeier, 2016). This period of development also coincides with enrollment in secondary school, which is a time of social and academic change for students.

Meanwhile, problematic internalizing and externalizing behaviors are also associated with poor academic performance. Externalizing behaviors can often predict future academic problems (Burt & Roisman, 2010; Vaillancourt, Brittain, McDougall, & Duku, 2013). Furthermore, severe internalizing behaviors are also associated with poorer performance, although mild to moderate internalizing is not linked to academic problems and indeed may be linked to better academic results (Riglin, Petrides, Frederickson, & Rice, 2014). One explanation is that poorer academic performance, as predicted by higher externalizing problems, leads to a lack of confidence and social withdrawal (Burt & Roisman, 2010; van der Ende et al., 2016).

2.2 The Strengths and Difficulties Questionnaire

The SDQ is a questionnaire with 25 short statements about specific behaviors. Separate forms exist for parent, teacher, and self-ratings. The respondent chooses whether each behavior statement is not true, somewhat true, or certainly true. The instrument is typically divided into either three or five subscales, which follow from theory on externalizing and internalizing behaviors (Goodman, Lamping, & Ploubidis, 2010). In the five-scale structure, there are five questions each scale, namely conduct problems, hyperactivity, emotional problems, peer problems, and prosocial behavior. In the three-scale structure, hyperactivity and conduct problems

are combined into a ten item externalizing scale. Similarly, emotional problems and peer problems make up the ten item internalizing scale, while the prosocial scale remains unchanged. An alternate formulation uses a single twenty-item total difficulties scale alongside the prosocial behavior scale. A list of all 25 items as well as the factor structure can be found in Table 2.1.

2.2.1 Classroom use of the SDQ.

Because of the important relationship between behavioral issues and academic development (e.g., van der Ende et al., 2016), the SDQ is commonly used by teachers and administrators to screen children for potential developmental issues and thus to also identify children who may encounter ongoing or future academic issues (Goodman, Renfrew, & Mullick, 2000). Gräf et al. (2019) recently examined in detail the relationship between school performance and the SDQ scales. Before children began attending school for the first time, they were given the SDQ, and then the relationship between SDQ results and first grade performance was analyzed. They found that externalizing problems predicted worse performance already during the first year of schooling, but this was not clearly evident for internalizing problems. Similarly, Mundy et al. (2017) found a relationship between higher SDQ scores and worse performance in 5th graders. In this case, a relationship was discovered between gender, high difficulties scores and subject-specific poor performance. In older children, high scores on SDQ problem scales relate to academic risks as well. Adolescents with lower socioeconomic status were found to have a significantly

Table 2.1

The Strengths and Difficulties Questionnaire

3-Factor Label	5-Factor Label	Number	Item Text
Internalizing	Emotional Problems	3	Often complains of headaches, stomach-aches or sickness
		8	Many worries or often seems worried
		13	Often unhappy, downhearted or tearful
		16	Nervous or clingy in new situations, easily loses confidence
		24	Many fears, easily scared
	Peer Problems	6	Rather solitary, tends to play alone
		11	<i>Has at least one good friend</i>
		14	<i>Generally liked by other children</i>
		19	Picked on or bullied by other children
		23	Gets on better with adults than with other children
		25	Gets on better with other children than with adults
Externalizing	Conduct Problems	5	Often has temper tantrums or hot tempers
		7	<i>Generally well behaved, usually does what adults request</i>
		12	Often fights with other children or bullies them
		18	Often lies or cheats
		22	Steals from home, school or elsewhere
	Hyperactivity	2	Restless, overactive, cannot sit still for Long
		10	Constantly fidgeting or squirming
		15	Easily distracted, concentration wanders
		21	<i>Thinks things out before acting</i>
		25	<i>Good attention span, sees tasks through to the end</i>
Prosocial Behavior	Prosocial Behavior	1	Considerate of other people's feelings
		4	Shares readily with other children, for examples toys, treats, pencils
		9	Helpful if someone is hurt, upset or feeling ill
		17	Kind to younger children
		20	Often volunteers to help others (parents, teachers, other children)

Note: Items in italics are reverse coded.

higher drop out rate in secondary school (Gross, Duncan, Kim, Alex Mason, & Haggerty, 2018).

2.2.2 Issues with the SDQ.

The SDQ's many language versions are widely used around the world. It is often used as a screening tool (e.g., Goodman et al., 2000) or to relate its scales to other to other latent variables, such as competency (e.g., Mundy et al., 2017), drop-out risk (e.g., Ingul & Nordahl, 2013), and other factors. However, before it can be used in all contexts, each version must be validated. Most validation studies consist of testing the factor structure of the instrument (e.g., Goodman et al., 2010; Ortuño-Sierra, Aritio-Solana, & Fonseca-Pedrero, 2018). Only a few have investigated its application to specific groups such as gender, different ages, immigration status and special education needs. One of these studies was Ortuño-Sierra, Fonseca-Pedrero et al. (2015) who provided a multi-cultural sample which demonstrated concerns that the instrument may lack invariance across cultures. Similarly, Hagquist (2007) showed that there are concerns for invariance across age groups within a Swedish sample. Others have found differing factor structures within language or culture specific versions of the SDQ (e.g., Di Riso et al., 2010; Ortuño-Sierra, Chocarro, Fonseca-Pedrero, Riba, & Muñiz, 2015; Ortuño-Sierra, Fonseca-Pedrero et al., 2015).

2.3 Social Behavior

The SDQ includes two subscales that examine social behavior directly, namely peer problems and prosocial behavior. Social behaviors can also affect competency and grades. According to Farrington et al. (2012) this may be a direct or an indirect

effect. First, teachers may be more likely to provide better evaluations and grades to students who engage in more positive social behavior. This represents a direct effect of a child's social behavior on his or her grades. However, teachers may also be more inclined to provide additional support or encouragement to these children as well, which describes an indirect effect of social behavior on not only the student's grades, but also likely their competency.

2.4 Inclusion in the Classroom

Children with special education needs face special challenges in the school system. They have higher levels of internalizing and externalizing problems (Turunen, Kiuru, Poskiparta, Niemi, & Nurmi, 2018; Weymeis, van Leeuwen, & Braet, 2018), are more likely to be excluded by their peers (Avramidis, 2010; Rose, Monda-Amaya, & Espelage, 2010; Schwab, Gebhardt, & Gasteiger-Klicpera, 2013) and are at risk for worse academic outcomes (Szumski & Karwowski, 2015; van Ryzin, Gravely, & Roseth, 2009). Classroom inclusion intends to incorporate such at risk children into the regular classroom. By incorporating children with SEN, it is expected to reduce feelings of exclusion (Grütter, Gasser, & Malti, 2017; Schwab, 2017) and boost academic performance.

Classroom inclusion and exclusion also relate to social behavior. Children who are excluded may face higher rates of peer problems, and children who engage in less prosocial behavior may similarly be more likely to be excluded (Ladd & Kochenderfer-Ladd, 2016). As social behavior may affect academic performance, this may only exacerbate the problems such children are already facing.

While there is some consensus that SDQ scores relate to academic performance, this relationship does not exist in isolation. The SDQ is also related to a number of variables such as inclusion in school and social behavior which can also affect academic performance. Children who are excluded typically have more externalizing and internalizing issues and may also have problematic social behaviors (DeRosier, Kupersmidt, & Patterson, 1994; Leary, 2001).

2.5 Perception of Inclusion Questionnaire

Although the SDQ has a teacher version, its scales are not specific to an academic setting. Rather, the SDQ focusses on psychosocial factors in a general context as rated by a teacher. The Perception of Inclusion Questionnaire (PIQ; Zurbriggen, Venetz, Schwab, & Hessels, 2017) provides an academically focused scale that assesses self-reported emotional inclusion, social inclusion, and academic self-concept. It was created by shortening the FDI (Fragebogen zur Erfassung der Dimensionen der Integration von Schülern; Haeberlin, Moser, Bless, & Klaghofer, 1989) to 12 items. It has multiple language versions and has been validated in several settings (Zurbriggen et al., 2017). The PIQ is detailed in Table 2.2.

Instead of examining problematic behaviors, the PIQ examines self-concept and inclusion within in an academic context. Previous studies suggest that children who have higher problem scores or related issues suffer socially in the school context and may have a lower academic self-concept (Elbaum & Vaughn, 2003; Szumski & Karwowski, 2015). The establishment of a relationship between problem scores on the SDQ and lower inclusion scores and academic self concept on the PIQ would

strengthen the theoretical relationship between social and emotional problems and academic difficulties.

Table 2.2

The Perception of Inclusion Questionnaire

Number	Factor	Text
1	Emotional	I like going to school.
4	Inclusion	<i>I have no desire to go to school.</i>
7		I like it in school.
10		School is fun.
2	Social	I have a lot of friends in my class.
5	Inclusion	I get along very well with my classmates.
8		<i>I feel alone in my class.</i>
11		I have very good relationships with my classmates.
3	Academic	I am a fast learner.
6	Self-Concept	I am able to solve very difficult exercises.
9		I do well in my schoolwork.
12		<i>Many things in school are too difficult for me.</i>

Note: Items in italics are reverse coded.

2.6 Research Questions

This research project represents a combined set of articles relating to the development, application, and validation of the SDQ and PIQ in the German academic setting. Which motivates the first main research question of the project: is the SDQ a valid measure? Critical to this question are examinations of the reported factor structure by previous research and measures of invariance across examined groups (i.e., gender, special education needs, and measurement point). The second, related, question is how do the scores of the SDQ vary over time? Related to this question is an examination of longitudinal development at various stages of

psychological development. It would be expected that changes in problem scores or prosocial behavior would correspond appropriately. An additional validation comes in the third main research question: how do the social subscales of the SDQ relate to academic achievement? Previous work has related social and other problems to poorer academic performance, but this can be particularly difficult to see for social problems. Finding a relationship between social problems and academic achievement would further validate the use of the SDQ in an academic setting. Finally, social problems relate to the last main research question: how do the SDQ and PIQ relate to each other? Specifically, do higher scores of inclusion and academic self-concept relate to lower problem scores on the SDQ and higher prosocial behavior scores? This would be expected if both instruments are measuring common or related variables. For instance, lower levels of inclusion are may be related to lower levels of prosocial behavior or higher levels of internalizing and externalizing problems.

2.7 Overview of the articles included in this Framework

The articles in this framework were designed to first explore the validity of the SDQ in the German school system and then relate the SDQ results to inclusion, academic self-concept, competency and grades. While many studies have examined individual relationships within this web of effects, these studies are designed to explore in more detail these interwoven relationships. The first article verified the factor structure and invariance of the SDQ across multiple measurement points.

After doing so, it longitudinally examined the development of the individual scales over each measurement point. The second article related the social scores (peer problems and prosocial behavior) to grades and competency in reading and math within a large-scale assessment. The third article examined the relationship between the SDQ and PIQ. First, it examined the invariance of both scales across multiple measurement points, gender, and special education needs. Next, it compared the levels of problem scores, prosocial behavior, emotional inclusion, social inclusion and academic concept within a correlation matrix. Finally, it examined how these scores develop over time and with relation to gender and special education needs.

3.0 Article One: “An assessment of measurement invariance in the 3- and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study”

3.1 Purpose and Motivation

In the first article, I examine the usefulness of the SDQ in longitudinal studies in a German sample. This was a necessary first step before future work with the SDQ. In particular, the psychometric properties of the scale required assessment. While the SDQ is used frequently around the world, frequent issues with specific language and regional versions are common (e.g., Ortuño-Sierra, Fonseca-Pedrero et al., 2015)). Moreover, the psychometric quality of the data over time was rarely assessed, and then only in specific regional version. Further, these assessments have typically been cross-sectional in design instead of longitudinal (e.g Di Riso et al., 2010; Hagquist, 2007).

One the psychometric properties across multiple measurement points can be verified, the SDQ could prove an excellent tool to track internalizing and externalizing across development.

3.2 Study Description

The study included data from a sample of students in a small city in northern Germany. The sample included all second year students enrolled in public elementary school at the first year of data collection (N = 433), and follow up data was collected for the same cohort for the next three years (i.e., 2nd-5th school years). Teachers rated their students on the 25-item German version of the SDQ. In this

school system, teachers follow their students as they progress to the next grade; this allowed for students to be rated by the same rater each year. This allowed for an unusually complete, longitudinal data set where the raters did not change from year to year.

I analyzed the SDQ data using the structural equation modeling (SEM) software Mplus 7.4 (Muthén & Muthén, 1998-2015). I assessed both the fits of both the 3-factor and 5-factor models in the dataset and further examined the invariance over the 4-measurement points. Finally, I examined the change in SDQ scores over the course of data collection with a repeated measures MANOVA.

3.3 Results and Discussion

Items 1 and 14 had to be removed from the analyses because they resulted in invalid models. Specifically, they perfectly predicted their associated latent variables resulting in invalid variance estimates. After removing these items, the overall fits for both the 3-factor and 5-factor models were considered to be good. Strong invariance was also upheld across all measurement points.

I expected to find good fits, but also expected that invariance would not be upheld. This is possibly because other studies typically examined invariance in a cross-sectional way or across cultures, instead of longitudinally. Regardless, strong invariance meant I could proceed to comparisons of sum scores over time.

MANOVA results indicated significant changes for all problem scores and prosocial behavior over time. Specifically, problem scores increased over the assessment period, while prosocial behavior decreased. This was true for both 3-

factor and 5-factor model. This was also expected based on theoretical models of development and academic performance (van der Ende et al., 2016) as well as findings from another German sample, albeit with an older age-group (Becker et al., 2015).

With the initial verification of the SDQ and an examination of development and change over time, the next step was to connect developmental issues to academic performance more directly.

3.4 Author Contributions

The data was collected under the supervision of Stefan Voß. This included basic data organization and input into digital files. The initial framework for the article was conceived by Markus Gebhardt. Jeffrey M. DeVries performed data checks, designed and performed the analyses, and wrote all sections of the article. Stefan Voß and Markus Gebhardt reviewed the article and provided comments for revision. Revisions were done by Jeffrey M. DeVries.

4.0 Article Two: “How Does Social Behavior Relate to Both Grades and Achievement Scores?”

4.1 Purpose and Motivation

The second article explores the relationship between scores on the two socially related subscales of the SDQ and academic achievement. I used both grades and competency in a longitudinal dataset as measures of academic achievement, and I explored how the relationship between social behavior and academic achievement varied between both achievement variables. Previous work has established that social behavior can impact academic achievement, but findings on this have often been inconsistent (Farrington et al., 2012). This could be because of the indirect nature of the effect of social behavior on achievement and due to using conflicting measures of academic achievement. For instance, grades may be affected by other variables that are weaker or absent when examining competency alone. These variables could include personality factors, motivation, and others (Farrington et al., 2012). A longitudinal study with a sufficiently large sample can explore the relationship of social behavior on grades alongside other moderating variables. In doing so, the link between social behavior and academic outcomes can be better understood.

4.2 Study Description

The study used data from the National Education Panel Study (NEPS) in Germany (Blossfeld, Roßbach, & von Maurice). Students in the cohort starting in year 5 were used (SC3; Buchholz & NEPS, National Educational Panel Study, 2017). Path

models were created using SEM in Mplus 7.4 (Muthén & Muthén, 1998-2015). The subscales for prosocial behavior and peer problems were modeled latently and then regressed onto both 5th and 7th year competencies and grades. Gender and parental education level were also regressed onto prosocial behavior, peer problems, and 5th and 7th year competency and grades. A similar model was created for reading where reading competency and German grades were substituted for math competency and grades. The model can be seen in Figure 4.1.

4.3 Results and Discussion

Prosocial behavior was found to significantly relate to better 5th year grades for math and German, but not 7th year grades. It did not relate to competency for either year. Peer problems related to worse grades in both 5th and 7th year for both math and German. Similarly, they related to worse math and reading competency in the 5th year, but they did not relate to any difference in competency in the 7th year.

Additional notable effects included significantly better grades and competencies for reading and math for children whose parents had a university degree. Furthermore, girls had better grades in both math and German in both years, while boys had better math competency in both years, but worse reading competency in year 7.

Overall, the results showed that prosocial behavior had a similar effect for both reading and math grades and did not relate to competency. Meanwhile, peer problems related to grades across both measurement points and competency in the 5th year. Peer problems are one of the two internalizing subscales of the SDQ. Thus

these results support the overall negative effect of internalizing problems on academic achievement. There is a clear, consistent relationship between this problem score from the SDQ and two separate, but related academic outcomes.

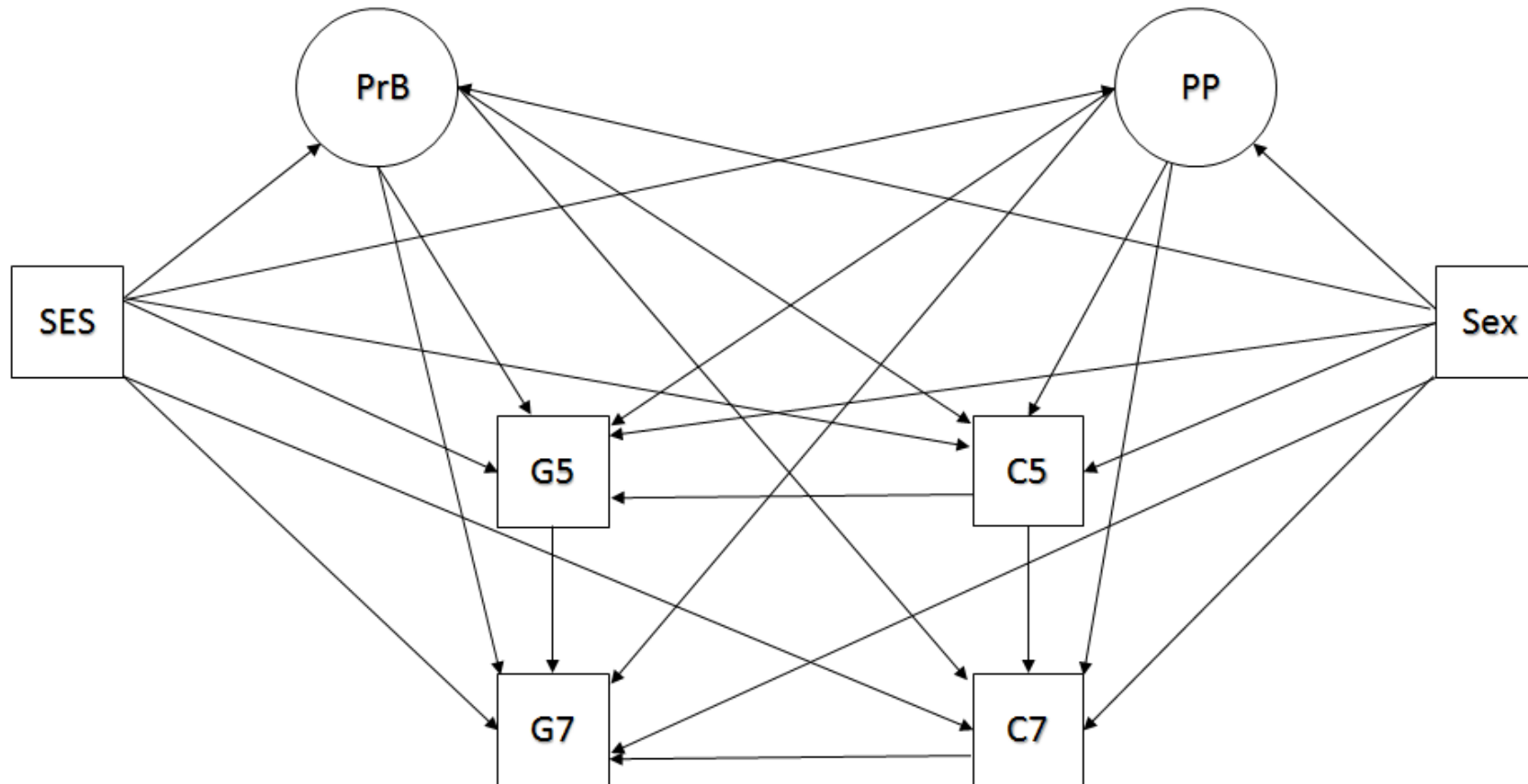
This may be related to poor social skills or social problems such as feelings of exclusion. The next and final study will examine in more detail the relationship between inclusion and exclusion in schools, academic self-concept and internalizing and externalizing problems in a school setting.

4.4 Author Contributions

Data was provided by NEPS. Markus Gebhardt provided writing oversight, feedback, and initial study design. Jeffrey M. DeVries served as primary author and data analyst. This included data exploration, analysis development, and final analysis design. Similarly, he wrote all sections of the paper, and implemented revisions suggested by Markus Gebhardt and Katharina Rathmann. Katharina Rathmann provided expertise on developing and implementing gender and socioeconomic status into the models, particularly within the framework of the NEPS databases.

Figure 4.1

Path Model for Social Behavior and Achievement



Note: SES stands for socioeconomic status. PrB stands for prosocial behavior. PP stands for peer problems. G5 and G7 stand for grades in year 5 and year 7 respectively. C5 and C7 stand for competency in year 5 and year 7 respectively.

5.0 Article Three: “Do learners with special education needs really feel included?”

Evidence from the Perception of Inclusion Questionnaire and Strengths and

Difficulties Questionnaire

5.1 Purpose and Motivation

This third and final paper explores the role of inclusion in another large sample of students. It connects personal characteristics, such as problem scores on the SDQ to academic self-concept and self-perceived social and emotional inclusion. This helped further detail the role of internalizing and externalizing problems in the academic realm. I further extended these analyses to an examination of SEN within inclusive schools.

The study examined the relationship between the subscales of SDQ and PIQ to further validate both instruments and explore the relationship between their associated latent constructs. I examined the structure and invariance of both instruments across grade level, SEN and gender. Once invariance was tested, I made compared the sum scores of these scales across grade level, SEN and gender.

5.2 Study Description

Participants were drawn from a complete sample (N = 407) of all 6th grade students in public early secondary schools in a single small city in northern Germany. The children were assessed at the end of their 6th year, and again at the end of their 7th year. The schools followed an established framework for inclusion according to local and regional guidelines (Voß et al., 2017). Within the sample were

48 children identified as having SEN. All students took the self-report version of both the SDQ and PIQ at both measurement points.

5.3 Results and Discussion

The 3-factor structure of the PIQ was assessed with a CFA in Mplus 7.4 (Muthén & Muthén, 1998-2015). Fit metrics indicated a good fit; however, initial fits of the 5-factor SDQ were insufficient. After minor modifications in line with previous work, the SDQ fits became acceptable. Across gender, grade level, and SENs, the PIQ demonstrated strong invariance. However, the SDQ only demonstrated strong invariance across SENs and grade level. Sufficient partial invariance (i.e., 20% freed loadings and intercepts; see (Dimitrov, 2010; Levine et al., 2003) could also not be demonstrated for gender on the SDQ. I concluded that it was acceptable to compare sum scores across groups, except for the SDQ across gender.

Next, I compared the subscales of each instrument to each other. It was found that higher prosocial behavior correlated with better feelings of social and emotional inclusion as well as academic self-concept. Meanwhile, higher scores for all four SDQ problem subscales related to worse levels of academic self-concept and social and emotional inclusion with the exception that hyperactivity did not relate significantly to feelings of social inclusion, and peer problems did not significantly relate to emotional inclusion.

Finally, I compared the sum scores for each factor based on grade level, gender, and SEN via a repeated measures MANOVA. Notably, I found that academic self-concept increased between the 6th and 7th year of school and that children with

SEN had a lower academic self-concept and feelings of emotional inclusion.

Furthermore, children with SEN had a higher level of conduct problems.

Importantly, school year interacted significantly with SEN on all three PIQ subscales. Children with SEN had an improved academic self-concept in year 7, but this was not at the level of children without SEN. They also had a higher level of emotional and social inclusion in year 7. In year 6, emotional and social inclusion were significantly below their peers without SEN, but by year 7, their feelings of inclusion were not different from their peers.

This third paper showed the relationship between the SDQ and another more academically focused instrument within a school environment. Children with SEN demonstrated a higher level of conduct problems, but not other problems, and prosocial behavior was also similar. The pattern of increased problems from paper 1 was also not replicated. This may have been due to an older sample (secondary school instead of primary) or a smaller sampling window (2 years instead of 4).

5.4 Author Contributions

Data were collected under the supervision of Stefan Voß. Jeffrey M. DeVries and Markus Gebhardt developed the initial idea and scope of the article. Jeffrey M. DeVries performed data analyses, corrections, and modeling. He wrote all sections of the article. Markus Gebhardt and Stefan Voß reviewed the article and recommended revisions which were written by Jeffrey M. DeVries.

6.0 General Discussion

6.1 Review of general results

My research indicates that the German version of the SDQ remains a valid tool for use in several academic contexts in Germany. It possess sufficient longitudinal invariance to make comparisons, and it relates to other academic achievement and school inclusion in predictable ways.

6.1.1 Validation of the SDQ

Numerous studies have reported various issues with different language versions of the SDQ (e.g., Di Riso et al., 2010; Hagquist, 2007; Ortuño-Sierra, Chocarro et al., 2015; Ortuño-Sierra, Fonseca-Pedrero et al., 2015). In this project, I identified several minor issues within the German version of the SDQ. Minor modifications were required in both the first and third article. To achieve acceptable fits within CFAs, two items were removed from the former case, and in the latter two items were allowed to covary. Although these modifications are in line with previous work, they may indicate areas of concern for future work. Despite these minor modifications, the 3- and 5-factor structure produces similar fits in the German version to other versions.

Further issues include the lack of strong measurement invariance in the third study. Specifically, strong invariance was not found for comparisons over time and across gender. While partial invariance was upheld for comparisons over time, this was not true for gender. It is therefore recommended that future researchers be very careful when making comparisons on SDQ scores. Further work into the functioning

of specific items across comparisons groups may be warranted. Nonetheless, in my samples, invariance was deemed sufficient to make comparisons over time and for children with and without SEN.

6.1.2 Longitudinal change in SDQ scores

Possibly conflicting results were found when comparing changes in SDQ over time in articles one and three. In article one, all problem scores increased and prosocial behavior decreased over the measurement intervals. However, no change over time was detected in article two. Based on the work of Becker et al. (2015), I expected SDQ scores to increase over time in both samples. Becker et al.'s sample included many different age groups and there was a general increase in problem scores over time; however, Becker et al. argued that the increase may be limited to specific at-risk individuals. It may be that the failure to find an increase in article three is related to a limited time window (two measurement points instead of four). It would also be useful to examine if the increases in SDQ problem scores affects all individuals or if it only affects a minority, as Becker et al. (2015) argue. Overall, I see some evidence for an increase in problem scores, but more work is necessary across different age groups and with multiple measurement points.

6.1.3 Relationship of the SDQ to Academic Achievement

In the second article, I examined the relationship between the social subscales of the SDQ and academic achievement. This study demonstrated that each subscale had a somewhat different relationship with both grades and competency. Specifically, prosocial behavior related to better math and German grades, but not to

competency. This was also restricted to year five grades. By year seven, prosocial behavior, as measured two years earlier, was not related to grades. Meanwhile, peer problems related to worse math and German grades for both year five and year seven, although the effect was stronger for year five grades. Peer problems also related to worse reading and math competency in year five, but not in year seven. This confirms the existence of the relationship between social behavior and academic achievement, but the nature of that relationship requires further examination.

It may be that the peer problems subscale is a stronger predictor of social difficulties than prosocial behavior, or that it is simply a stronger predictor of academic difficulties in general. Either possibility could account for the differential relationship to year five and seven achievement. It is also possible that scores on prosocial behavior and peer problems would change significantly between these years and thus mask the relationship. However, this is complicated by conflicting evidence about change in SDQ scores from articles one and three.

It is also unclear the underlying mechanism of the relationship between social behavior and achievement. Farrington et al. (2012) argued that this relationship was an indirect one and might be characterized by varying types of social interactions with the teacher and classmates. This could have an effect on both direct instruction from the teacher and group work activities. Alternatively, teachers may give better grades and more attention to students with greater social skills, or social behavior may correlate to some other personality factor that relates more strongly to grades. More work is necessary to understand this relationship.

6.1.4 Inclusion and the SDQ

A clear relationship was found between SDQ responses and the perception of inclusion in the third article. With two notable exceptions, children with high problem scores felt less emotionally and socially included at school, and they had a lower academic self-concept. The exceptions were that hyperactivity scores did not relate to social inclusion and peer problems scores did not relate to emotional inclusion. The exceptions may indicate that hyperactive children might suffer in the classroom while not being excluded by their peers, and children with peer problems may be socially excluded, but still like school itself. More work should examine these exceptions. Nonetheless, the very consistent correlations between perception of inclusion and SDQ problem scores indicate that children with more internalizing and externalizing problems generally feel less included and have a lower academic self-concept.

Article three also examined differences in SDQ and PIQ scales for children with and without SEN. Children with SEN had a lower academic self concept and perception of emotional inclusion, although there was no difference in their perception of social inclusion. The only difference between these groups on the SDQ was that children with SEN had more conduct problems. Meanwhile, children with SEN had an increase in all three scores between years six and seven. In the case of social and emotional inclusion, they were on the same level as children without SEN. This may be indicative of an ongoing process where children with SEN feel excluded at first, but feel more included over time within an inclusive school.

6.2 Limitations & Future Work

There are several limitations of these articles. First of all, all work included only German children and adolescents, and moreover, both articles one and three included samples from only one city within Germany. Comparison studies are needed to relate these findings to those in other countries and regions. These studies only relate the SDQ to specific academic skills and to a single other instrument. More work connected both the SDQ to other instruments and outcome variables is also necessary. Lastly, while I included two longitudinal studies, all of this work relies on correlations. More work examining the nature of these correlations and the possibility or likelihood of other related variables is necessary. Some of these variables were included in the second study (gender and parental education), but other variables, such as household income, migration status, specific types of SEN, developmental variables, intelligence and others, must also be identified and examined. These may include other variables about the home situation, such as parental style, variables relating to the individual, such as motivation, learning styles, and reasoning ability, as well as teacher related variables, such as teaching style and experience and qualities of an inclusive classroom.

6.3 Conclusion

The SDQ is in general a valid instrument for use within German schools, but attention must be paid to invariance when comparing two or more groups. The social subscales of the SDQ show a clear relationship to academic achievement, but the nature of this relationship requires further study. The relationship between the SDQ

and PIQ indicates that children with higher levels internalizing and externalizing behaviors often feel less included in school, and that children with SEN feel more included over time. These variables are of particular interest to educators because of students with higher levels internalizing and externalizing problems perform worse academically. Identification of such problems is the first step for teachers, administrators, and parents who wish to assist students in this way. The SDQ is recommended for teachers, researchers, and other practitioners to use the SDQ for this task. Although scores on either should not be taken as definitive evidence of a developmental or other problems, they may allow practitioners additional evidence of the sort of problems their learners are facing, thus allowing them to more directly craft lessons and support for their students.

Given the relative simplicity to administer and the wealth of information about potential student issues, the psychometric quality, and the connections to other academic variables, the SDQ can be used within schools to identify children who demonstrate potentially problematic internalizing and externalizing behaviors. After which interventions to counteract these behaviors may be introduced, which may have positive learning outcomes for the student.

7.0 References

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, 80(7), 1–37.
<https://doi.org/10.1037/h0093906>
- Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85(6), 1275–1301. <https://doi.org/10.1037/0033-2909.85.6.1275>
- Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/Externalizing Problems: Review and Recommendations for Clinical and Research Applications. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(8), 647–656. <https://doi.org/10.1016/j.jaac.2016.05.012>
- Avramidis, E. (2010). Social relationships of pupils with special educational needs in the mainstream primary class: Peer group membership and peer-assessed social behaviour. *European Journal of Special Needs Education*, 25(4), 413–429.
<https://doi.org/10.1080/08856257.2010.513550>
- Beauchaine, T. P., & Crowell, S. E. (Eds.). (2018). *The Oxford handbook of emotion dysregulation*. Oxford: Oxford University Press.
- Becker, A., Rothenberger, A., Sohn, A., Ravens-Sieberer, U., & Klasen, F. (2015). Six years ahead: A longitudinal analysis regarding course and predictive value of the Strengths and Difficulties Questionnaire (SDQ) in children and adolescents. *European Child & Adolescent Psychiatry*, 24(6), 715–725.
<https://doi.org/10.1007/s00787-014-0640-x>

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). [Special issue]. *Zeitschrift für Erziehungswissenschaft*, 14.

Buchholz, S., & NEPS, National Educational Panel Study. (2017). *NEPS-Startkohorte 3: Klasse 5 (SC3 7.0.0)*.

Burt, K. B., & Roisman, G. I. (2010). Competence and psychopathology: Cascade effects in the NICHD Study of Early Child Care and Youth Development. *Development and Psychopathology*, 22(3), 557–567.

<https://doi.org/10.1017/S0954579410000271>

DeRosier, M. E., Kupersmidt, J. B., & Patterson, C. J. (1994). Children's Academic and Behavioral Adjustment as a Function of the Chronicity and Proximity of Peer Rejection. *Child Development*, 65(6), 1799–1813. <https://doi.org/10.1111/j.1467-8624.1994.tb00850.x>

DeVries, J. M., Gebhardt, M., & Voß, S. (2017). An assessment of measurement invariance in the 3- and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study. *Personality and Individual Differences*, 119, 1–6. <https://doi.org/10.1016/j.paid.2017.06.026>

Di Riso, D., Salcuni, S., Chessa, D., Raudino, A., Lis, A., & Altoè, G. (2010). The Strengths and Difficulties Questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Differences*, 49(6), 570–575. <https://doi.org/10.1016/j.paid.2010.05.005>

- 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. <https://doi.org/10.1177/0748175610373459>
- Elbaum, B., & Vaughn, S. (2003). For which students with learning disabilities are self-concept interventions effective? *Journal of Learning Disabilities*, 36(2), 101-8; discussion 149-50. <https://doi.org/10.1177/002221940303600203>
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191. <https://doi.org/10.1007/s10802-010-9434-x>
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(11), 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015>
- Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European child & adolescent psychiatry*, 9(2), 129–134. <https://doi.org/10.1007/s007870050008>

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note.

Journal of Child Psychology and Psychiatry, 38(5), 581–586.

<https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>

Gräf, C., Hoffmann, I., Diefenbach, C., König, J., Schmidt, M. F., Schnick-Vollmer,

K., . . . Urschitz, M. S. (2019). Mental health problems and school performance in

first graders: Results of the prospective cohort study ikidS. *European Child &*

Adolescent Psychiatry. Advance online publication. [https://doi.org/10.1007/s00787-](https://doi.org/10.1007/s00787-019-01296-7)

[019-01296-7](https://doi.org/10.1007/s00787-019-01296-7)

Gross, T. J., Duncan, J., Kim, S. Y., Alex Mason, W., & Haggerty, K. P. (2018).

Predicting School Suspension Risk from Eighth Through Tenth Grade Using the

Strengths and Difficulties Questionnaire. *Contemporary School Psychology*, 4(3), 100.

<https://doi.org/10.1007/s40688-018-00215-y>

Grütter, J., Gasser, L., & Malti, T. (2017). The role of cross-group friendship and

emotions in adolescents' attitudes towards inclusion. *Research in Developmental*

Disabilities, 62, 137–147. <https://doi.org/10.1016/j.ridd.2017.01.004>

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 ; mit einem Heft zur theoretischen und praktischen Einführung, einem Testbogen

zur Beurteilung des sozialen, emotionalen und leistungsmotivationalen Integriertseins von

Schülern des 4. bis 6. Schuljahres in ihre Schulklasse und Auswertungsschablonen

[Questionnaire to assess the dimensions of Integration of pupils; FDI 4-6; an assessment of

the social, emotional, and motivational Integration of 4th-6th grade students, including a theoretical and practical introduction and evaluation templates]. Bern: Haupt.

Hagquist, C. (2007). The psychometric properties of the self-reported SDQ – An analysis of Swedish data based on the Rasch model. *Personality and Individual Differences, 43*(5), 1289–1301. <https://doi.org/10.1016/j.paid.2007.03.022>

Hostinar, C. E., & Cicchetti, D. (2018). Emotion Dysregulation and Internalizing Spectrum Disorders. In T. P. Beauchaine & S. E. Crowell (Eds.), *The Oxford handbook of emotion dysregulation*. Oxford: Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780190689285.013.18>

Ingul, J. M., & Nordahl, H. M. (2013). Anxiety as a risk factor for school absenteeism: What differentiates anxious school attenders from non-attenders? *Annals of General Psychiatry, 12*(1), 25. <https://doi.org/10.1186/1744-859X-12-25>

Ladd, G. W., & Kochenderfer-Ladd, B. (2016). Research in Educational Psychology: Social Exclusion in School. In P. Riva & J. Eck (Eds.), *Social Exclusion: Psychological Approaches to Understanding and Reducing Its Impact* (pp. 109–132). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33033-4_6

Leary, M. R. (Ed.). (2001). *Interpersonal rejection*. New York: Oxford University Press.
Retrieved from <http://site.ebrary.com/lib/academiccompletetitles/home.action>

Levine, D. W., Kaplan, R. M., Kripke, D. F., Bowen, D. J., Naughton, M. J., & Shumaker, S. A. (2003). Factor structure and measurement invariance of the Women's Health Initiative Insomnia Rating Scale. *Psychological Assessment, 15*(2), 123–136. <https://doi.org/10.1037/1040-3590.15.2.123>

Mundy, L. K., Canterford, L., Tucker, D., Bayer, J., Romaniuk, H., Sawyer, S., . . .

Patton, G. (2017). Academic Performance in Primary School Children With Common Emotional and Behavioral Problems. *The Journal of School Health, 87*(8), 593–601. <https://doi.org/10.1111/josh.12531>

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide*. Los Angeles, CA.

Oland, A. A., & Shaw, D. S. (2005). Pure versus co-occurring externalizing and internalizing symptoms in children: The potential role of socio-developmental milestones. *Clinical Child and Family Psychology Review, 8*(4), 247–270.

<https://doi.org/10.1007/s10567-005-8808-z>

Ortuño-Sierra, J., Aritio-Solana, R., & Fonseca-Pedrero, E. (2018). Mental health difficulties in children and adolescents: The study of the SDQ in the Spanish National Health Survey 2011-2012. *Psychiatry Research, 259*, 236–242.

<https://doi.org/10.1016/j.psychres.2017.10.025>

Ortuño-Sierra, J., Chocarro, E., Fonseca-Pedrero, E., Riba, S. S. I., & Muñoz, J. (2015).

The assessment of emotional and Behavioural problems: Internal structure of The Strengths and Difficulties Questionnaire. *International Journal of Clinical and Health Psychology : IJCHP, 15*(3), 265–273. <https://doi.org/10.1016/j.ijchp.2015.05.005>

Ortuño-Sierra, J., Fonseca-Pedrero, E., Aritio-Solana, R., Velasco, A. M., Luis, E. C. de,

Schumann, G., . . . Lawrence, C. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European Child & Adolescent Psychiatry, 24*(12), 1523–1534. <https://doi.org/10.1007/s00787-015-0729->

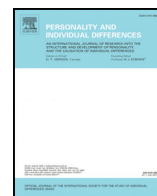
- Riglin, L., Petrides, K. V., Frederickson, N., & Rice, F. (2014). The relationship between emotional problems and subsequent school attainment: A meta-analysis. *Journal of Adolescence, 37*(4), 335–346.
<https://doi.org/10.1016/j.adolescence.2014.02.010>
- Rose, C. A., Monda-Amaya, L. E., & Espelage, D. L. (2010). Bullying Perpetration and Victimization in Special Education: A Review of the Literature. *Remedial and Special Education, 32*(2), 114–130. <https://doi.org/10.1177/0741932510361247>
- Schwab, S. (2017). The impact of contact on students' attitudes towards peers with disabilities. *Research in Developmental Disabilities, 62*, 160–165.
<https://doi.org/10.1016/j.ridd.2017.01.015>
- Schwab, S., Gebhardt, M., & Gasteiger-Klicpera, B. (2013). Predictors of social inclusion of students with and without SEN in integrated settings. *Hrvatska Revija Za Rehabilitacijska Istraživanja, 49*(SUPPL.), 106–114.
- Soto-Sanz, V., Castellví, P., Piqueras, J. A., Rodríguez-Marín, J., Rodríguez-Jiménez, T., Miranda-Mendizábal, A., . . . Alonso, J. (2019). Internalizing and externalizing symptoms and suicidal behaviour in young people: A systematic review and meta-analysis of longitudinal studies. *Acta Psychiatrica Scandinavica, 140*(1), 5–19.
<https://doi.org/10.1111/acps.13036>
- 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. <https://doi.org/10.1016/j.lindif.2015.08.037>

- Theunissen, M. H. C., Wolff, M. S. de, & Reijneveld, S. A. (2019). The Strengths and Difficulties Questionnaire Self-Report: A Valid Instrument for the Identification of Emotional and Behavioral Problems. *Academic Pediatrics, 19*(4), 471–476.
<https://doi.org/10.1016/j.acap.2018.12.008>
- Turunen, T., Kiuru, N., Poskiparta, E., Niemi, P., & Nurmi, J.-E. (2018). Word Reading Skills and Externalizing and Internalizing Problems from Grade 1 to Grade 2—Developmental Trajectories and Bullying Involvement in Grade 3. *Scientific Studies of Reading, 23*(2), 161–177.
<https://doi.org/10.1080/10888438.2018.1497036>
- Vaillancourt, T., Brittain, H. L., McDougall, P., & Duku, E. (2013). Longitudinal links between childhood peer victimization, internalizing and externalizing problems, and academic functioning: Developmental cascades. *Journal of Abnormal Child Psychology, 41*(8), 1203–1215. <https://doi.org/10.1007/s10802-013-9781-5>
- Van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2016). The bidirectional pathways between internalizing and externalizing problems and academic performance from 6 to 18 years. *Development and Psychopathology, 28*(3), 855–867.
<https://doi.org/10.1017/S0954579416000353>
- Van Ryzin, M. J., Gravely, A. A., & Roseth, C. J. (2009). Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being. *Journal of Youth and Adolescence, 38*(1), 1–12. <https://doi.org/10.1007/s10964-007-9257-4>

Voß, S., Hauer, A., Blumenthal, Y., Mahlau, K., Sikora, S., & Hartke, B. (2017). *Zum Leistungs- und Entwicklungsstand inklusiv beschulter Schülerinnen und Schüler mit (sonder-)pädagogischen Förderbedarfen auf der Insel Rügen nach sechs Schulbesuchsjahren [Evaluation and state of development for inclusion of students with special education needs on the island of Rügen after six years of schooling]*. Retrieved from https://www.rim.uni-rostock.de/fileadmin/uni-rostock/Alle_PHF/RIM/Downloads/RIM-Evaluationsbericht-MZP7_Internet.pdf

Weymeis, H., van Leeuwen, K., & Braet, C. (2018). Adaptive emotion regulation, academic performance and internalising problems in Flemish children with special educational needs: A pilot study. *European Journal of Special Needs Education, 34*(1), 124–135. <https://doi.org/10.1080/08856257.2017.1421601>

Zurbriggen, C. L. A., Venetz, M., Schwab, S., & Hessels, M. G. P. (2017). A Psychometric Analysis of the Student Version of the Perceptions of Inclusion Questionnaire (PIQ). *European Journal of Psychological Assessment, 1–9*. <https://doi.org/10.1027/1015-5759/a000443>



An assessment of measurement invariance in the 3- and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study



Jeffrey M. DeVries^{a,*}, Markus Gebhardt^a, Stefan Voß^b

^a Technische Universität Dortmund, Germany

^b Universität Rostock, Germany

ARTICLE INFO

Article history:

Received 27 February 2017

Received in revised form 8 June 2017

Accepted 16 June 2017

Available online xxxx

Keywords:

Strengths and Difficulties Questionnaire

SDQ

Structural equation modeling

SEM

Measurement invariance

Internalizing

Externalizing

ABSTRACT

The Strengths and Difficulties Questionnaire (SDQ) is an internationally popular tool for assessing of psychological and psychosocial development; however, recent research has suggested it lacks measurement invariance in several situations. We expand on this discussion by examining the measurement variance of its two popular models across longitudinal measurements. We examined the fits of both models, conducted tests of invariance across time, and measured the change over time for each factor. We found that both models provide good fits and, furthermore, that strong measurement invariance was upheld across all measurement points for both models. Additionally, we found that problem scores increased and prosocial behavior scores decreased over time. We conclude that both models provide good fits of the data and have measurement invariance over time, but that more research is required regarding its measurement invariance across cultures.

© 2017 Published by Elsevier Ltd.

1. Introduction

As a short and flexible questionnaire, the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997, 2001) is a popular tool to assess potential developmental issues. However, there remains a lack of consensus to the interpretation and psychometric properties of the SDQ (Di Riso et al., 2010; Hagquist, 2007; McCrory & Layte, 2012). Additionally, there exists uncertainty about proper interpretation of the SDQ scales, which is complicated by items where measurement invariance is lacking. Furthermore, measurement invariance has not been investigated via longitudinal comparisons. Because the SDQ is used as a screening tool for potential developmental problems (Goodman, 1997), longitudinal data is an important component of validation. This paper will provide an important validation of the SDQ using longitudinal comparisons and afterwards, examine longitudinal development in its key psychological scales.

The SDQ (Goodman, 1997, 2001) is used in many countries and dozens of different languages to assess student mental health, screen for mental disorders, identify other risk factors, and predict academic performance (Kersten et al., 2016; Woerner et al., 2004). It contains 25 behavior-based-items, which are ranked on a 3-factor Likert scale (“Not True”, “Somewhat True”, and “Certainly True”). The test has

three versions, which are either parent-rated, teacher-rated, or self-rated. Regardless of version, there are two major models that interpret results, the 5-factor and 3-factor models (Goodman, Lamping, & Ploubidis, 2010). The 5-factor model contains the factors of emotional problems, conduct problems, hyperactivity, peer problems, and prosocial behavior. Alternatively, the 3-factor model uses the factors of internalizing problems, externalizing problems, and prosocial behavior. Internalizing problems are a combination of emotional and peer problems, and similarly externalizing problems include conduct problems and hyperactivity.

Research on both the 3- and 5- factor models indicates that both provide an acceptable fit (Di Riso et al., 2010; Goodman et al., 2010). However, there is evidence that individual items may lack measurement invariance based on either cultural specific factors (Ortuno-Sierra et al., 2015) or time factors (Hagquist, 2007).

1.1. Theoretical basis of both models

The 3-factor model is based on Achenbach's (1966) theory regarding internalizing and externalizing behaviors, where behaviors are categorized based on the locus of their target. Externalizing behaviors include behaviors such as fighting, yelling, stealing, while internalizing behaviors may include fearfulness, anxiety, and withdrawal (Achenbach, Ivanova, Rescorla, Turner, & Althoff, 2016). Both behavior types have

* Corresponding author.

E-mail address: jeffrey.devries@tu-dortmund.de (J.M. DeVries).

been associated with different academic and psychological problems in developing children and adolescents.

Longitudinal analyses show that externalizing problems are apparent before academic problems arise (Burt & Roisman, 2010; Vaillancourt, Brittain, McDougall, & Duku, 2013; van der Ende, Verhulst, & Tiemeier, 2016). However, internalizing behaviors have been linked to improved academic performance (Vaillancourt et al., 2013). Vaillancourt et al. (2013) argue that internalizing children may be insulated from externalizing problems. One possibility is that low to moderate rates of internalizing behavior are associated with increased performance, while high rates of internalizing behavior decrease performance. This is supported by Riglin, Petrides, Frederickson, and Rice's (2014) meta-analysis, where only severe internalizing behaviors have an overall effect on academic performance. Additionally, it is noteworthy that both types of behaviors can coincide within the same child (Oland & Shaw, 2005). van der Ende et al. (2016) presented a longitudinal model in which externalizing problems lead to poor academic performance, which then led to feelings of incompetence and withdrawal. Thus, externalizing problems may lead to later coinciding internalizing problems.

Meanwhile, the 5-factor model may offer a more precise description of specific behaviors and difficulties that make up externalizing and internalizing behavior. It has been shown to be a useful screening tool for early diagnoses of multiple childhood disorders, such as the two different externalizing disorders of conduct disorder and hyperkinetic disorder (Goodman, Renfrew, & Mullick, 2000).

Both models of the SDQ use sum scores of their latent factors to assess behavior levels and to match these levels with the above described patterns of psychological development (Goodman et al., 2010). However, a prerequisite of making comparisons with sum scores is an assessment of the measurement invariance of the model (van de Schoot, Lugtig, & Hox, 2012).

1.2. SDQ, age and invariant measurement

One recent longitudinal study of psychological development examined the diagnostic features of the SDQ in German adolescents over six-years (Becker, Rothenberger, Sohn, Ravens-Sieberer, & Klasen, 2015). Across four measurement points, there was little volatility for adolescents who were rated normal, but there was moderately greater volatility for students rated borderline or abnormal. This coincided with the finding that emotional problems increased from ages 9 to 16, and hyperactivity decreased.

However, to interpret these results as meaningful changes over time, the SDQ must show measurement invariance over time, and there is evidence that the SDQ lacks measurement invariance in cross-sectional studies. Hagquist (2007) examined invariant measurement across time for multiple cohorts, finding several items that lacked measurement invariance. This matches the findings of additional studies that have reported on the SDQ's reliability, model fits and measurement invariance across different age groups in Denmark (Niclasen, Skovgaard, Andersen, Somhøvd, & Obel, 2013) and across different cultures (Ortuno-Sierra et al., 2015).

These findings stress the importance of verifying the fits and measurement invariance as a prerequisite to making meaningful inferences about the development of psychological scales over time.

1.3. The present study

The present study will expand on the work investigating the invariance of the SDQ in two key ways. First, we will present teacher ratings to add to previous work focusing on self-report versions of the SDQ (e.g., Hagquist, 2007; Ortuno-Sierra et al., 2015). Second, we will present longitudinal data with each child rated by the same teacher at each measurement point. Then, we assess the invariance for both the 3-factor and 5-factor models of the German SDQ. Once strong

measurement invariance is confirmed, we will then examine changes in the scales of the SDQ over time.

1.4. Hypotheses

Our hypotheses follow the order of our analyses. First, we verified the model fits in our present dataset meet the standards of previous studies with the SDQ. Second, we assess the measurement invariance of our models. Third, we compare the sum scores across our four measurement points.

1) Both models will fit the data well. The SDQ has been extensively tested and validated by SEM in both the 3-factor and 5-factor models (Di Riso et al., 2010; Goodman et al., 2010; Ortuno-Sierra et al., 2015). Nevertheless, fit-values of both models must be established before assessing measurement invariance.

2) Strong invariance will not be upheld across all measurement points in both models. While weak (non-scalar) invariance has been upheld across cultures and over time in several studies, strong invariance has not been found for all items (Hagquist, 2007; Niclasen et al., 2013; Ortuno-Sierra et al., 2015). These items are marked in Table 1.

3) There will be increases in the four problem scores and decreases in the prosocial behavior score over time. Higher ratings on both internalizing and externalizing problems over time were found in another German sample from Becker et al. (2015), which matches the overall developmental trajectories of internalizing and externalizing problems (e.g., Vaillancourt et al., 2013; van der Ende et al., 2016). Corresponding to increased internalizing and externalizing patterns, we expect lower prosocial behavior (e.g., (Marryat, Thompson, Minnis, & Wilson, 2014).

Table 1
The 3- and 5-factor models of the SDQ.

3-Factor latent variable	5-Factor latent variable	Item	Item in English
Internalizing	Emotional problems	3	Often complains of headaches, stomach-aches or sickness
		8	Many worries or often seems worried
		13	Often unhappy, downhearted or tearful
		16	Nervous or clingy in new situations, easily loses confidence
		24	Many fears, easily scared
		6	Rather solitary, tends to play alone
	Peer problems	11	<i>Has at least one good friend</i>
		14	Generally liked by other children
		19	<i>Picked on or bullied by other children</i>
		23	Gets on better with adults than with other children
Externalizing	Conduct problems	5	Often has temper tantrums or hot tempers
		7	Generally well behaved, usually does what adults request
		12	Often fights with other children or bullies them
		18	<i>Often lies or cheats</i>
		22	Steals from home, school or elsewhere
	Hyperactivity	2	Restless, overactive, cannot sit still for long
		10	Constantly fidgeting or squirming
		15	Easily distracted, concentration wanders
		21	Thinks things out before acting
		25	Good attention span, sees tasks through to the end
Prosocial behavior	Prosocial behavior	1	Considerate of other people's feelings
		4	Shares readily with other children, for examples toys, treats, pencils
		9	Helpful if someone is hurt, upset or feeling ill
		17	Kind to younger children
		20	Often volunteers to help others (parents, teachers, other children)

Note: Items in bold are the items that lacked strong measurement invariance across cultures (Ortuno-Sierra et al., 2015). Italicized items are items where measurement invariance across ages was not found due to disordered thresholds (Hagquist, 2007).

2. Method

2.1. Participants

Participants were a cohort of all children ($N = 433$, 224 girls) enrolled from 2010 to 2014 in all public elementary schools in one small city in northeastern Germany. This included 11 different schools with one school teaching in a Montessori style. Each student was measured once at the end of the school year. The total participants at each measurement point varied due to normal migration of families, absence caused by illness, and parental consent ($N_{t1} = 289$, $N_{t2} = 342$, $N_{t3} = 370$, $N_{t4} = 375$). Their ages varied consistently for each measurement point over three years ($M_{t1} = 7.8$ years, $SD_{t1} = 0.4$ years; $M_{t4} = 10.8$, $SD_{t4} = 0.4$).

2.2. Data collection

Students were rated by their teachers on the German version SDQ at the end of every school year from 2011 to 2014. Students remained with the same teacher each year, resulting in each student being rated by the same person each year.

2.3. Instrument

The German language teacher SDQ was used. Responses were scored as 0, 1, or 2, with higher values representing a higher score on the scale. Table 1 lists all 25 items.

2.4. Analyses

Table 1 shows the 3-factor and the 5-factor models of the SDQ. Each latent variable correlates to the other latent variables at the same measurement point. We calculated a separate confirmatory factor analysis for each time point.

Analyses were conducted with Mplus 7.4 (Muthén & Muthén, 1998–2015). Fit measures were calculated with the weighted least squares with mean and variance adjusted (WLSMV), which is considered the most appropriate estimator for ordinal items with interval latent variables (Flora & Curran, 2004; Muthén, Du Toit, & Spisic, 1997). Each school was treated as a separate cluster of data in a complex analysis to control for the effect of a nested sample. We report fit measures including the root-mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker Lewis Index (TLI). Acceptable fits include $RMSEA < 0.08$, $CFI > 0.90$, and $TLI > 0.90$. Good fits include $RMSEA < 0.05$, $CFI > 0.95$, and $TLI > 0.95$ (Brown, 2006).

In measurement points 2, 3, and 4, analyses for the 5-factor model produced unacceptable SEM solutions. Items 1 and 14 were perfectly linearly correlated with their latent variables, likely due to missing values, and sample size. To verify, we repeated the analysis using only participants with complete data ($n = 180$), but encountered the same problem. We continued with the full subject pool on the future SEM analyses and we excluded both items in all models. This is consistent with minor modifications made by past researchers to obtain better fits (e. g., Hagquist, 2007).

In order to compare the means of latent variables, scalar (strong) measurement invariance is required (Steinmetz, 2013; van de Schoot et al., 2012). In the scalar model, both factor loadings and intercepts are kept constant in each group. To test for strong measurement invariance, we compared the scalar model to the configural model in Mplus (Muthén & Muthén, 1998–2015). We reported chi-square tests, RMSEA, CFI, and TLI values for both scalar and configural models.

Finally, to test hypothesis 3, we performed repeated measures ANOVAs on the sum-scores of the 5-factor model of the SDQ. For this analysis, we only included participants with full data at every measurement point, resulting in a reduced sample size ($n = 180$).

3. Results

3.1. Hypothesis 1: SEM fits

A brief summary of the model fits can be found in Table 2. Both models met our criteria for a good fit, with RMSEA below 0.05 and CFI and TLI above 0.95. Although the CFI values are below threshold at time point 2 for the 3-factor model, the RMSEA is still below threshold. Additionally, factor loadings are presented in Table 3; all loadings were significant for both models at each measurement point.

We conclude from the fit-metrics and significant factor loadings that both models provide a good fit to our data. While at some measurement points the 5-factor model outperforms the 3-factor model, overall both models maintain a good fit. Therefore, we deemed it appropriate to proceed with assessing measurement invariance.

3.2. Hypothesis 2: measurement invariance

Model fits for our invariance tests are presented in Table 4. We found both models met the standards for scalar invariance, 3-factor $\chi^2(120) = 142.70$, $p = 0.078$, 5-factor $\chi^2(108) = 123.88$, $p = 0.141$. According to Cheung and Rensvold (2002), a $\Delta CFI = 0.01$ is a sufficient threshold for the model with fewer constraints. When applying this criterion, we find that values for both models are below threshold, $\Delta CFI(3\text{-factor}) = 0.002$, $\Delta CFI(5\text{-factor}) = 0.003$. Contradicting hypothesis 2, we conclude that scalar invariance is upheld for both models, so it was unnecessary to further investigate specific items by individually freeing them from constraints (Byrne, Shavelson, & Muthén, 1989; van de Schoot et al., 2012).

3.3. Hypothesis 3: development of scores over time

Table 5 shows the sum scores for each category both models and how many students were categorized as borderline or abnormal on each factor in the 5-factor model. The repeated measures ANOVA failed tests of sphericity for each factor, all $\chi^2(5) > 21$, $p < 0.001$. Therefore, the Greenhouse-Geisser correction was used.

3.3.1. Emotional problems

Emotional problems varied significantly across measurement points, Greenhouse-Geisser $F(2.68, 479.00) = 7.252$, $p < 0.001$, partial $\eta^2 = 0.04$. Table 5 shows an increase over time and more students meeting criteria for borderline and abnormal levels.

3.3.2. Peer problems

Peer problems ratings varied significantly across measurement points, Greenhouse-Geisser $F(2.65, 473.95) = 18.331$, $p < 0.001$, partial $\eta^2 = 0.09$. As with emotional problems, Table 5 shows increased values over time, and more students meeting criteria for borderline and abnormal levels.

3.3.3. Conduct problems

Conduct problems also varied significantly across measurement points, Greenhouse-Geisser $F(2.70, 482.43) = 11.977$, $p < 0.001$, partial

Table 2
Model fits.

		df	RMSEA	90% CI RMSEA	CFI	TLI
Time 1	3-Factor	227	0.048	0.039–0.056	0.941	0.934
	5-Factor	220	0.040	0.030–0.049	0.959	0.953
Time 2	3-Factor	227	0.033	0.029–0.046	0.957	0.952
	5-Factor	220	0.033	0.023–0.041	0.969	0.965
Time 3	3-Factor	227	0.043	0.035–0.050	0.958	0.953
	5-Factor	220	0.037	0.028–0.045	0.970	0.966
Time 4	3-Factor	227	0.038	0.030–0.046	0.974	0.971
	5-Factor	220	0.033	0.024–0.042	0.981	0.978

Table 3
Standardized factor loadings with SE.

3F latent variable	5F latent variable	Item	Time 1		Time 2		Time 3		Time 4	
			3F	5F	3F	5F	3F	5F	3F	5F
Internalizing	Emotional problems	3	0.57 (0.09)***	0.63 (0.09)***	0.64 (0.07)***	0.66 (0.07)***	0.58 (0.04)***	0.64 (0.05)***	0.64 (0.04)***	0.66 (0.04)***
		8	0.91 (0.02)***	0.92 (0.03)***	0.85 (0.04)***	0.86 (0.03)***	0.81 (0.03)***	0.84 (0.03)***	0.83 (0.04)***	0.87 (0.04)***
		13	0.93 (0.02)***	0.97 (0.02)***	0.90 (0.03)***	0.92 (0.03)***	0.77 (0.04)***	0.87 (0.04)***	0.88 (0.04)***	0.89 (0.05)***
		16	0.76 (0.06)***	0.80 (0.06)***	0.86 (0.04)***	0.90 (0.04)***	0.82 (0.05)***	0.93 (0.07)***	0.85 (0.03)***	0.91 (0.04)***
	Peer problems	24	0.70 (0.06)***	0.75 (0.06)***	0.79 (0.04)***	0.79 (0.04)***	0.67 (0.06)***	0.71 (0.06)***	0.82 (0.02)***	0.83 (0.02)***
		6	0.53 (0.10)***	0.54 (0.10)***	0.53 (0.04)***	0.57 (0.04)***	0.56 (0.06)***	0.58 (0.05)***	0.58 (0.05)***	0.67 (0.06)***
		11	0.87 (0.04)***	0.90 (0.06)***	0.77 (0.08)***	0.85 (0.07)***	0.77 (0.03)***	0.87 (0.03)***	0.66 (0.09)***	0.76 (0.09)***
		19	0.73 (0.06)***	0.70 (0.07)***	0.73 (0.06)***	0.84 (0.06)***	0.91 (0.03)***	0.92 (0.04)***	0.74 (0.05)***	0.86 (0.08)***
		23	0.35 (0.13)**	0.34 (0.13)**	0.52 (0.14)***	0.57 (0.15)***	0.54 (0.08)***	0.58 (0.07)***	0.52 (0.08)***	0.61 (0.09)***
		22	0.72 (0.08)***	0.71 (0.08)***	0.74 (0.05)***	0.78 (0.06)***	0.62 (0.11)***	0.63 (0.11)***	0.79 (0.08)***	0.81 (0.09)***
Externalizing	Conduct problems	5	0.80 (0.02)***	0.82 (0.03)***	0.85 (0.04)***	0.94 (0.04)***	0.86 (0.03)***	0.90 (0.03)***	0.89 (0.03)***	0.88 (0.03)***
		7	0.74 (0.02)***	0.80 (0.02)***	0.72 (0.05)***	0.81 (0.05)***	0.83 (0.02)***	0.87 (0.02)***	0.83 (0.02)***	0.90 (0.01)***
		12	0.99 (0.03)***	0.98 (0.03)***	0.81 (0.03)***	0.86 (0.03)***	0.90 (0.01)***	0.91 (0.01)***	0.85 (0.01)***	0.88 (0.02)***
		18	0.74 (0.05)***	0.74 (0.05)***	0.75 (0.03)***	0.79 (0.03)***	0.75 (0.02)***	0.79 (0.03)***	0.79 (0.02)***	0.83 (0.02)***
	Hyper-activity	2	0.84 (0.03)***	0.88 (0.04)***	0.93 (0.01)***	0.94 (0.01)***	0.91 (0.02)***	0.93 (0.05)***	0.89 (0.03)***	0.92 (0.03)***
		10	0.84 (0.03)***	0.87 (0.03)***	0.94 (0.01)***	0.96 (0.01)***	0.92 (0.02)***	0.92 (0.02)***	0.86 (0.03)***	0.89 (0.04)***
		15	0.85 (0.03)***	0.87 (0.03)***	0.89 (0.01)***	0.90 (0.01)***	0.82 (0.02)***	0.84 (0.02)***	0.93 (0.01)***	0.93 (0.02)***
		21	0.86 (0.03)***	0.90 (0.03)***	0.78 (0.03)***	0.81 (0.03)***	0.79 (0.02)***	0.79 (0.02)***	0.84 (0.03)***	0.86 (0.03)***
		25	0.62 (0.05)***	0.70 (0.04)***	0.91 (0.02)***	0.92 (0.02)***	0.79 (0.01)***	0.82 (0.01)***	0.85 (0.02)***	0.87 (0.02)***
		20	0.89 (0.08)***	0.88 (0.02)***	0.81 (0.04)***	0.81 (0.04)***	0.88 (0.04)***	0.88 (0.03)***	0.91 (0.04)***	0.90 (0.04)***
Prosocial behavior	Prosocial behavior	4	0.75 (0.04)***	0.77 (0.04)***	0.87 (0.03)***	0.87 (0.03)***	0.82 (0.02)***	0.83 (0.02)***	0.84 (0.02)***	0.84 (0.02)***
		9	0.93 (0.06)***	0.92 (0.06)***	0.85 (0.05)***	0.85 (0.05)***	0.83 (0.03)***	0.80 (0.03)***	0.95 (0.01)***	0.96 (0.01)***
		17	0.52 (0.08)***	0.53 (0.08)***	0.74 (0.04)***	0.74 (0.04)***	0.74 (0.03)***	0.75 (0.03)***	0.84 (0.03)***	0.85 (0.03)***
		20	0.89 (0.08)***	0.88 (0.02)***	0.81 (0.04)***	0.81 (0.04)***	0.88 (0.04)***	0.88 (0.03)***	0.91 (0.04)***	0.90 (0.04)***

3F refers to the 3-factor model, and 5F refers to the 5-factor model.
 ** p < 0.01.
 *** p < 0.001.

$\eta^2 = 0.06$. These changes are reflected in Table 5, where we can see an increase in mean values and number meeting the criteria for borderline or abnormal diagnoses.

3.3.4. Hyperactivity

Hyperactivity values varied across time, Greenhouse–Geisser $F(2.76, 479.00) = 3.15, p < 0.05$, partial $\eta^2 = 0.02$. Table 5 shows a slight increase over time, along with a corresponding increase in proportion of students rated as abnormal on the scale.

3.3.5. Prosocial behavior

Prosocial behavior ratings varied across time, Greenhouse–Geisser $F(2.54, 453.71) = 12.391, p < 0.001$ partial $\eta^2 = 0.06$. As seen in Table 5, prosocial behavior declined over time, which corresponded to increases in the number of students rated borderline or abnormal by the scale.

Overall, effect sizes were small to moderate, but significant. All five factors grow significantly worse over the course of the four measurement points. This consistent finding supports the hypothesis that internalization and externalization scores increase over time, while prosocial scores decrease over time.

4. Discussion

This paper provided a set of novel analyses of longitudinal SDQ data with each student rated by the same teacher at each measurement point. Using this data, we compared both the 3- and 5-factor models for the SDQ based on both model fit and measurement invariance and then examined the sample-level changes in SDQ latent variables over time.

The overall goodness of fit of both models is consistent with findings of other researchers regarding the SDQ (Niclasen et al., 2013; Ortuno-Sierra et al., 2015), and the conclusions by Goodman et al. (2010) that both models fit similarly. While we did find that the 5-factor model produced a better fit, the differences between both models were quite small. Therefore, we concluded that both models were in line with previous research and proceeded to examine the measurement invariance of both models.

In that line, we were unable to support previous findings of a lack of strong measurement invariance in the SDQ over time. In cross-sectional studies in Sweden (Hagquist, 2007) and Denmark (Niclasen et al., 2013), strong measurement invariance was not upheld over age groups for all items. One explanation could be the test version. We used the teacher version; however, Hagquist's (2007) used the self-report version. While Niclasen et al.'s (2013) data included both teacher and

Table 4
Tests of measurement invariance.

	χ^2	df	p-Value	RMSEA	RMSEA 90% CI	CFI	TLI	Δ CFI
3-Factor								
Configural	1452.38	908		0.042	0.038–0.046	0.963	0.959	
Scalar	1525.32	1028	0.078	0.038	0.034–0.041	0.966	0.967	0.003
5-Factor								
Configural	1260.37	880		0.035	0.031–0.040	0.974	0.970	
Scalar	1333.68	988	0.141	0.032	0.027–0.036	0.976	0.976	0.002

Note: Configural and scalar models were not significantly different from each other for both 3- and 5-factor models, indicating measurement invariance is upheld across all 4 measurement points. Δ CFI refers to the change from the configural to the scalar models.

Table 5
Development of SDQ sum-scores over time.

	Time 1	Time 2	Time 3	Time 4
Age – mean (SD)	7.7 (0.3)	8.8 (0.4)	9.8 (0.4)	10.8 (0.4)
Emotional problems***				
Mean (SD)	1.03 (1.73)	0.84 (1.37)	1.28 (1.86)	1.41 (1.88)
% borderline	3.3%	2.2%	2.2%	4.4%
% abnormal	3.3%	1.1%	5.6%	4.4%
Peer problems***				
Mean (SD)	1.29 (1.44)	1.17 (1.44)	1.32 (1.71)	2.02 (2.02)
% borderline	4.4%	6.7%	6.7%	11.1%
% abnormal	3.3%	2.2%	6.7%	7.8%
Conduct problems***				
Mean (SD)	0.82 (1.45)	0.94 (1.45)	1.24 (1.91)	1.40 (1.85)
% borderline	6.1%	8.9%	11.7%	10.6%
% abnormal	6.1%	5.6%	12.2%	12.8%
Hyperactivity**				
Mean (SD)	2.91 (2.58)	2.63 (2.63)	2.86 (2.67)	3.06 (2.88)
% borderline	6.7%	3.9%	5.0%	5.0%
% abnormal	8.9%	9.4%	10.0%	11.7%
Prosocial behavior***				
Mean (SD)	8.42 (1.89)	8.32 (1.90)	8.29 (1.99)	7.61 (2.37)
% borderline	5.0%	4.4%	8.9%	8.9%
% abnormal	5.0%	3.3%	3.3%	10.6%
Internalizing***				
Mean (SD)	2.33 (2.68)	2.01 (2.41)	2.60 (3.17)	3.42 (3.43)
Externalizing***				
Mean (SD)	3.73 (3.61)	3.57 (3.77)	4.11 (4.27)	4.46 (4.46)

All tests include Greenhouse-Geisser corrections. SDQ scoring does not contain borderline and abnormal thresholds for internalizing and externalizing scales. All statistics and tests presented in this table are only for the 180 subjects who had data available at every point (listwise deletion).

** $p < 0.01$.

*** $p < 0.001$.

parent data, they lacked teacher data for some measurement points. An additional possibility is that our study included a German sample. Other studies have indicated a lack of invariance across cultural groups (e.g., Ortuno-Sierra et al., 2015), so the test may demonstrate invariance over development but not across cultures. Alternatively, an interaction of these factors is possible.

Lastly, the changes in sum-scores over time are in line with previous work that indicated a growth in these factors over time. In van der Ende et al.'s (2016) bidirectional model, internalizing and externalizing problems create a feedback loop, where such problems interfere with academic performance, which in turn leads to an increase in externalizing and internalizing behaviors. We found an overall increase in these behaviors from ages 7 to 11, which mirrors findings in similar work involving adolescents (Becker et al., 2015). While we found measurement invariance for our items, it is possible other external factors affected our results.

There were several important limitations in our study. First, we were only able to measure children of ages 7 to 11 in a non-clinical sample. A clinical sample, or one with a broader age range, may produce different conclusions about model selection, invariance and changes over-time. Second, while we were able to provide consistency in our ratings by having each child rated by the same rater, it is possible changes in a teacher's rating standard affected the observed changes. Third, we were unable to include two items in our SEM models or invariance tests. It is possible that the inclusion of these items would change these results. Lastly, we did not investigate measurement invariance over time and across cultures. It is possible there is an interaction between these factors that our sample missed. Future research should include a larger and more multicultural sample to overcome these limitations and investigate the possibility of a culture by time interaction.

5. Conclusion

Both the 3-factor and 5-factor interpretations of the teacher SDQ meet standards for good fits and possess strong measurement

invariance, longitudinally. Furthermore, sample-level changes in SDQ scores fit the predictions of theory regarding internalization and externalizing. Therefore, we conclude that the psychometric properties including measurement invariance over time of the full version of the teacher-based SDQ are sufficient for its continued use measuring internalizing and externalizing changes in a monocultural, longitudinal context. However, more research is required regarding inconsistent measurement invariance findings from previous multicultural cross-sectional studies and for the different test versions.

References

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, 80(7), 1–37. <http://dx.doi.org/10.1037/h0093906>.
- Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/externalizing problems: Review and recommendations for clinical and research applications. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(8), 647–656. <http://dx.doi.org/10.1016/j.jaac.2016.05.012>.
- Becker, A., Rothenberger, A., Sohn, A., Ravens-Sieberer, U., & Klases, F. (2015). Six years ahead: A longitudinal analysis regarding course and predictive value of the Strengths and Difficulties Questionnaire (SDQ) in children and adolescents. *European Child & Adolescent Psychiatry*, 24(6), 715–725. <http://dx.doi.org/10.1007/s00787-014-0640-x>.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research. Methodology in the social sciences*. New York, London: Guilford Press (Retrieved from <http://www.loc.gov/catdir/enhancements/fy0629/2006001103-b.html>).
- Burt, K. B., & Roisman, G. I. (2010). Competence and psychopathology: Cascade effects in the NICHD study of early child care and youth development. *Development and Psychopathology*, 22(3), 557–567. <http://dx.doi.org/10.1017/S0954579410000271>.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. <http://dx.doi.org/10.1037/0033-2909.105.3.456>.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5.
- Di Riso, D., Salcuni, S., Chessa, D., Raudino, A., Lis, A., & Altoè, G. (2010). The Strengths and Difficulties Questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Differences*, 49(6), 570–575. <http://dx.doi.org/10.1016/j.paid.2010.05.005>.
- van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2016). The bidirectional pathways between internalizing and externalizing problems and academic performance from 6 to 18 years. *Development and Psychopathology*, 28(3), 855–867. <http://dx.doi.org/10.1017/S0954579416000353>.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. <http://dx.doi.org/10.1037/1082-989X.9.4.466>.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191. <http://dx.doi.org/10.1007/s10802-010-9434-x>.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(11), 1337–1345. <http://dx.doi.org/10.1097/00004583-200111000-00015>.
- Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, 9(2), 129–134.
- Hagquist, C. (2007). The psychometric properties of the self-reported SDQ – An analysis of Swedish data based on the Rasch model. *Personality and Individual Differences*, 43(5), 1289–1301. <http://dx.doi.org/10.1016/j.paid.2007.03.022>.
- Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, 40(1), 64–75. <http://dx.doi.org/10.1177/0165025415570647>.
- Marryat, L., Thompson, L., Minnis, H., & Wilson, P. (2014). Associations between social isolation, pro-social behaviour and emotional development in preschool aged children: A population based survey of kindergarten staff. *BMC Psychology*, 2(1), 44. <http://dx.doi.org/10.1186/s40359-014-0044-1>.
- McCorry, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences*, 52(8), 882–887. <http://dx.doi.org/10.1016/j.paid.2012.02.011>.
- Muthén, B. O., Du Toit, S., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Retrieved from https://www.statmodel.com/download/Article_075.pdf.
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (Seventh edition) (Los Angeles, CA).

- Niclasen, J., Skovgaard, A. M., Andersen, A. -M. N., Somhovd, M. J., & Obel, C. (2013). A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): A large scale cohort study. *Journal of Abnormal Child Psychology*, 41(3), 355–365. <http://dx.doi.org/10.1007/s10802-012-9683-y>.
- Oland, A. A., & Shaw, D. S. (2005). Pure versus co-occurring externalizing and internalizing symptoms in children: The potential role of socio-developmental milestones. *Clinical Child and Family Psychology Review*, 8(4), 247–270. <http://dx.doi.org/10.1007/s10567-005-8808-z>.
- Ortuno-Sierra, J., Fonseca-Pedrero, E., Aritio-Solana, R., Velasco, A. M., de Luis, E. C., Schumann, G., ... Lawrence, C. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European Child & Adolescent Psychiatry*, 24(12), 1523–1534. <http://dx.doi.org/10.1007/s00787-015-0729-x>.
- Riglin, L., Petrides, K. V., Frederickson, N., & Rice, F. (2014). The relationship between emotional problems and subsequent school attainment: A meta-analysis. *Journal of Adolescence*, 37(4), 335–346. <http://dx.doi.org/10.1016/j.adolescence.2014.02.010>.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *The European Journal of Developmental Psychology*, 9(4), 486–492. <http://dx.doi.org/10.1080/17405629.2012.686740>.
- Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology*, 9(1), 1–12. <http://dx.doi.org/10.1027/1614-2241/a000049>.
- Vaillancourt, T., Brittain, H. L., McDougall, P., & Duku, E. (2013). Longitudinal links between childhood peer victimization, internalizing and externalizing problems, and academic functioning: Developmental cascades. *Journal of Abnormal Child Psychology*, 41(8), 1203–1215. <http://dx.doi.org/10.1007/s10802-013-9781-5>.
- Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiari, G., Dagalarrondo, P., ... Tannock, R. (2004). The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child & Adolescent Psychiatry*, 13(Suppl. 2), II47–II54. <http://dx.doi.org/10.1007/s00787-004-2008-0>.



How Does Social Behavior Relate to Both Grades and Achievement Scores?

Jeffrey M. DeVries*, Katharina Rathmann and Markus Gebhardt

Faculty of Rehabilitation Sciences, Technische Universität Dortmund, Dortmund, Germany

Prosocial behavior and peer problems are an important correlate of academic development; however, these effects vary by achievement measures and social behaviors. In this paper, we examined data from the German National Education Panel Study (NEPS), and we use structural equation modeling (SEM) to model the effects of prosocial behavior and peer problems on grades and competencies for both math ($n = 3,310$) and reading ($n = 3,308$) in grades 5 and 7. Our models account for the moderating effect of both gender and socioeconomic status (SES) as determined by parental education. We conclude that social behaviors relate to grades more strongly than competencies, that peer problems relate more strongly to achievement than prosocial behavior, and that the relationship is weaker in later grades. We discuss the implication that grades and achievement tests are not interchangeable measures for educators and researchers.

OPEN ACCESS

Edited by:

Puri Checa,
Universidad de Granada, Spain

Reviewed by:

Angela Jocelyn Fawcett,
Swansea University, United Kingdom
Jesús Nicasio García Sánchez,
Universidad de León, Spain

*Correspondence:

Jeffrey M. DeVries
jeffrey.devries@tu-dortmund.de

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 07 February 2018

Accepted: 14 May 2018

Published: 04 June 2018

Citation:

DeVries JM, Rathmann K and
Gebhardt M (2018) How Does Social
Behavior Relate to Both Grades and
Achievement Scores?
Front. Psychol. 9:857.
doi: 10.3389/fpsyg.2018.00857

Keywords: prosocial behavior, peer problems, grades, competency, large-scale assessment, structural equation modeling, academic achievement

INTRODUCTION

Academic progress can be measured in multiple ways including grades and achievement scores, but these methods are not interchangeable. Grades are more strongly connected to multiple noncognitive factors, including social behaviors, than achievement tests (Borghans et al., 2011; Farrington et al., 2012; Lechner et al., 2017). Although social behaviors are an indirect predictor, they can broadly predict future academic success (Durlak et al., 2010). However, due to their indirect nature, sufficiently large-scale studies are required to discern the differential relationship social behaviors have with both grades and achievement scores. The National Education Panel Study (NEPS; Blossfeld et al., 2011) is a large-scale longitudinal study of multiple cohorts of German students, which gives a unique opportunity to examine such relationships. In this paper, we model the relationship between social behaviors (specifically prosocial behavior and peer problems), competency, and grades with data from NEPS, in order to unravel which academic measures (grades vs. achievement scores) correlate with social behavior.

Social Behavior and Academic Achievement

Within the social-emotional learning framework, social behaviors support the social medium of learning (Vygotsky, 1978; Slavin, 1995, 2014; Baroody et al., 2016). Farrington et al. (2012) list social behaviors as one of five critical noncognitive factors that predict success beyond school. Two specific types of behaviors can be linked to academic achievement: prosocial behavior and peer problems. These two behaviors have been linked to various academic skills such as study

habits, and classroom behavior, and peer interactions, which in turn affect academic performance. Wentzel (1993, 1998) has repeatedly found a strong link between prosocial behavior and academic achievement. More recently, Gerbino et al. (2018) analyzed data from an Italian large-scale assessment. They demonstrated that prosocial behavior remained a significant predictor of grades even after accounting for other variables such as personality factors and IQ. Relatedly, Lewis et al.'s (2017) large-scale twin study indicated that prosocial behavior substantially improved predications based on genetics and environmental characteristics. Similarly, peer problems also correlate to lower achievement (Wentzel and Caldwell, 1997), and Malecki and Elliot (2002) found that poor social skills indicated worse performance on achievement tests. More recently, Askell-Williams and Lawson (2015) showed that children with peer problems were more likely to have lower academic motivation as well as other school-related difficulties.

Nonetheless, some inconsistent results remain. Adams et al. (1999) found that after accounting for hyperactivity, conduct problems, and emotional problems, neither peer problems nor prosocial behavior related to math achievement test results; however, prosocial behavior remained related to reading achievement test results. This contrasts with Gerbino et al. (2018) results which indicated that prosocial behavior remains a significant correlate of overall grades after accounting for multiple other factors.

Grades vs. Achievement Tests

One factor that could help explain such discrepancies is the use of grades vs. achievement tests to measure academic achievement. For instance, many educators include behavior measures in their grading (Cross and Frary, 1999), and grades have been shown to reflect numerous personality factors in addition to academic competence (Borghans et al., 2011; Andrei et al., 2015; Lechner et al., 2017; Gerbino et al., 2018). For example, Lockl et al. (2017) found that theory of mind in kindergarten predicted grades in grade 1 and 2, but they did not examine any connection to achievement test scores. Moreover, theory of mind represents a specific aspect of social development, and more research examining peer problems and prosocial behavior is needed. Despite this, large-scale studies examining both grades and achievement testing alongside social behavior are rare.

Moderating Variables

Among others, two key moderating variables in these studies have been socio-economic status (SES) and gender. Children of higher SES tend to show fewer social problems and more prosocial behavior (Letourneau et al., 2013). They have higher levels of inclusion at school (Veland et al., 2015), receive better grades (Lekholm and Cliffordson, 2008), and perform better on other achievement measures (Sirin, 2005). Furthermore, lower SES children engage in more prosocial behavior (Piff and Robinson, 2017), but they are also at higher risk of developing social problems (Bradley and Corwyn, 2002). Additionally, well established differences have been found in developmental trajectories for boys and girls for prosocial behavior and peer problems (Card et al., 2008; Chaplin and Aldao, 2013), as well as

in both math and reading achievement (Robinson and Lubienski, 2011). It is therefore important to consider both gender and SES as important moderators when examining achievement and social behavior.

The Present Study

This study investigates the differential effects of prosocial behavior and peer problems on both grades and achievement tests. We examine both math and reading achievement measures in a longitudinal, large-scale assessment, and account for both gender and socioeconomic status (SES). The use of large-scale panel data is important because the effects of social behavior are predicted to be important, but indirect (Farrington et al., 2012). Because such indirect effects are a particularly difficult hurdle when predicting effects of different strengths, we use the NEPS database (Blossfeld et al., 2011), which includes data from a large-scale German longitudinal survey with enough participants to model all necessary variables.

Based on the role of social skills as a noncognitive factor in learning (see Farrington et al., 2012), we expect that more desirable social behavior will correlate to both better grades and better competencies in reading and math. In a recent similar study, internalizing problems were shown to have a detrimental effect on achievement outcomes of secondary students (Deighton et al., 2018). However, because grades are a better reflection of noncognitive factors in learning, our first prediction is that grades will be more impacted by social behaviors than competency (see Borghans et al., 2011; Lechner et al., 2017). Furthermore, both gender and SES are well-known moderators of achievement and social behavior. Therefore, our second prediction is males will do better on math measures while females will do better on reading measures, and that students with higher SES will outperform those with lower SES on both measures. In a similar analysis, (Gerbino et al., 2018) showed that effects of social behaviors on grades remained after accounting for moderating personality factors. Therefore, our final predication is that the effects of prosocial behavior and peer problems will remain after accounting for gender and SES as determined by parental education.

METHODS

Data and Participants

All data came from the NEPS database (Blossfeld et al., 2011), which contains multiple large representative cohorts of German students. NEPS data are collected each year from selected students, teachers, parents, and administrators. We focused on NEPS cohort 3, which began in grade 5. We used data from waves 1 (grade 5, October 2010–January 2011), 2 (grade 6, October 2011–January 2012), and 3 (grade 7, October 2012–January 2013). All participants with data on any of the key variables were included in our models. Because of small differences in who took the reading and math competency NEPS tests and in who reported their grades for German and math, the number of participants varied slightly between both datasets. We provide an overview of the participants in **Table 1**.

TABLE 1 | Participant information.

	Math model (<i>n</i> = 3310)	Reading model (<i>n</i> = 3308)
GENDER (PERCENT)		
Male	51.6%	50.6%
Female	48.2%	49.4%
AGE (MEAN, SD)		
Years	12.0 (0.8)	12.0 (0.8)
PARENTAL EDUCATION (PERCENT)		
Basic	14.0%	14.0%
Vocational	56.0%	56.0%
University	30.0%	30.0%
SCHOOL TYPE (PERCENT)		
Secondary—Hauptschule	7.6%	7.6%
Secondary—Realschule	22.1%	22.2%
Secondary—Gymnasium	52.2%	52.4%
Other	18.1%	17.8%

Parental Education was determined by CASMIN.

Data Collection

We focused on a small subset of the collected data for our models: math competency, math grades, SDQ scores for the subscales of peer problems and prosocial behavior, gender, and parental education level.

Competency Measures

We used the uncorrected weighted maximum likelihood estimates (WLE) from grades 5 and 7 in the NEPS dataset for both math and reading competency. Analyses by the NEPS team confirmed unidimensionality, reliability, and measurement invariance of these estimates across gender, books in household, and migration background (Haberhorn et al., 2012; Krannich et al., 2017). Math and Reading competency were assessed in waves one and three (grades five and seven).

Grades

Self-reported math and German whole-year grades were used for grades 5 and 7. In the German school system, grades are ordered from 1 to 5, with lower scores representing better grades (1 = very good, 2 = good, 3 = satisfactory, 4 = sufficient, 5 = failing).

Prosocial Behavior and Peer Problems

The prosocial behavior and peer problems subdimensions of the Strengths and Difficulties Questionnaire (SDQ) were used to assess social behavior in wave two. The SDQ is a frequently used questionnaire to assess psychological characteristics of children (Goodman, 1997; Goodman et al., 2010) and has been demonstrated to meet basic psychometric properties for longitudinal analyses in German samples (DeVries et al., 2017). The other three SDQ subscales were unavailable in the NEPS database for this time period.

Socioeconomic Status (SES)

In parent interviews in wave one, a parent responded about his or her own educational attainment as well as his or

her partner's attainment. Responses were rated based on the Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) scale (Brauns et al., 2003). The scale was reduced to three basic categories: low (no secondary degree, or secondary degree with basic vocational training), intermediate (advanced vocational training or vocational postsecondary school), and high (university level or higher). Only the higher rating from either parent was used for each child.

Analysis

We analyzed the data with structural equation modeling (SEM). Separate models were calculated for math and reading. A confirmatory factor analysis was conducted for each model with prosocial behavior and peer problems treated as latent variables calculated from individual items from relevant SDQ subscales. Additionally as depicted in **Figures 1, 2**, gender, parental education, grades (5th and 7th year), and competency were regressed onto each other and the latent variables. Mplus was used for all SEM analyses (Muthén and Muthén, 1998–2017), and an example of our Mplus instruction file is available in the Appendix. Estimations were performed using robust maximum likelihood estimation (MLR), and we report root mean square error of approximation (RMSEA), comparative fit index (CFI) and square root mean residual (SRMR). Acceptable fits included RMSEA < 0.08, CFI > 0.90, and SRMR < 0.10, and good fits included RMSEA < 0.05, CFI > 0.95, and SRMR < 0.08 (Hu and Bentler, 1998).

RESULTS

Model Fits

Math

Overall, the math model produced a good fit of the data, RMSEA = 0.036 (90% CI = 0.033–0.039), CFI = 0.93, and SRMR = 0.040. While the CFI falls below our threshold of 0.05, it remains in the acceptable range. Despite this, the RMSEA and SRMR are well below the threshold for a good fit. We therefore concluded we had a good fit.

Reading

Similarly, the reading model provided a good fit, RMSEA = 0.044 (90% CI = 0.038–0.049), CFI = 0.92, and SRMR = 0.046. As in the math model, the CFI was below threshold for a good fit, but was in the range of acceptable fits. Given the good values for the RMSEA and SRMR, we concluded that the fit was good.

Reliability and Factor Loadings for the Latent Factors

Cronbach's α for peer problems was 0.60, and for prosocial behavior was 0.71, while McDonald's total ω for peer problems was 0.61 and for prosocial behavior was 0.72. Factor loadings for both the math and reading models can be seen in **Table 2**. They were significant at $p < 0.001$, and ranged between 0.38 at and 0.69. While Cronbach's α and McDonald's ω for the prosocial peer problems were low, overall the measures performed similarly to values from the meta-analysis conducted by Stone et al. (2010). Given the acceptable fit values and overall good model fits, we

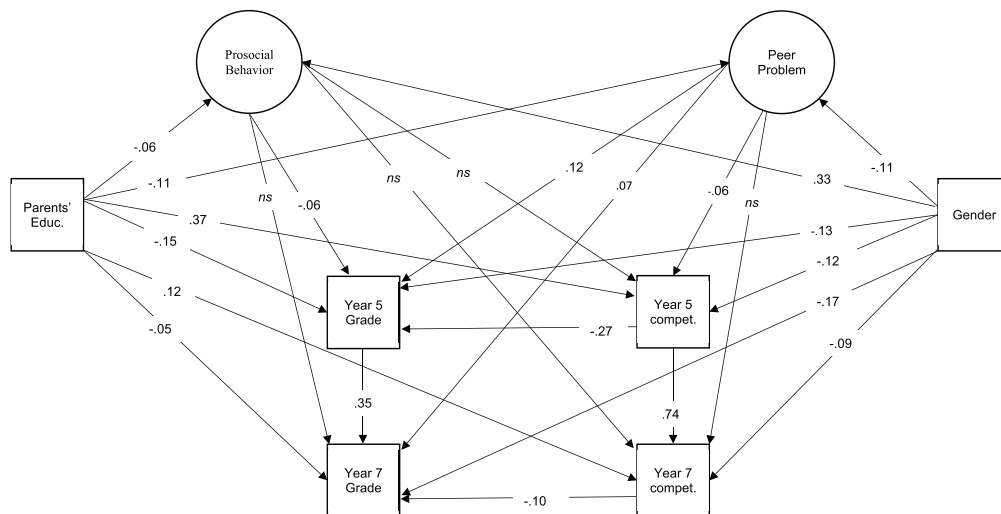


FIGURE 1 | Math Model with Significant Path Loadings. Parents' Educ. refers to parental education level as determined by CASMIN. Compet. refers to uncorrected WLE reported from NEPS competency assessments. Grades refer to final grade in the previous year. Factor loadings of SDQ items for the Prosocial Behavior and Peer Problems scales can be seen in **Table 2**.

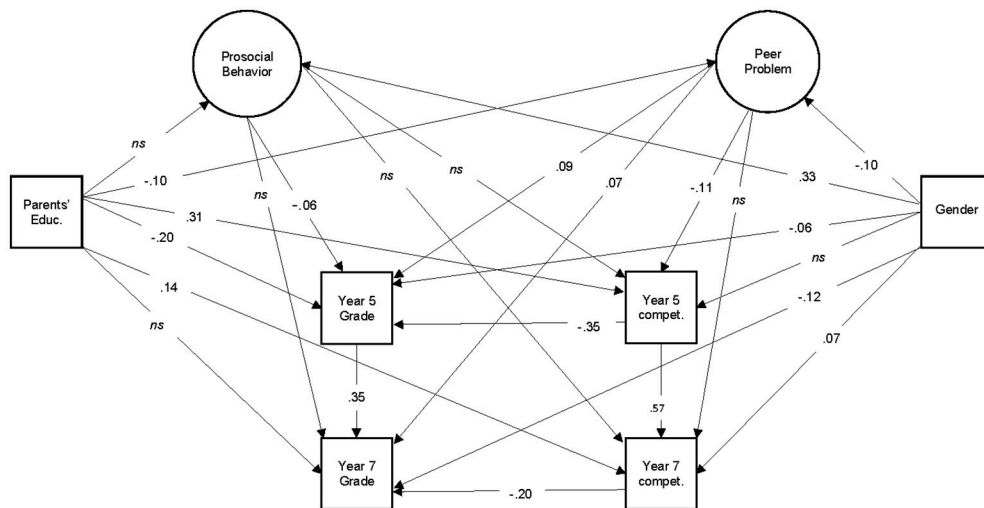


FIGURE 2 | Reading Model with Significant Path Loadings. Parents' Educ. refers to parental education level as determined by CASMIN. Compet. refers to uncorrected WLE reported from NEPS competency assessments. Grades refer to final grade in the previous year. Factor loadings of SDQ items for the Prosocial Behavior and Peer Problems scales can be seen in **Table 2**.

conclude the models fit the data reasonably well and provided sufficient reliability.

General Findings of Prosocial Behavior and Peer Problems

The standardized path loadings are reported in **Figures 1, 2**. Prosocial behavior only related to both math and reading grades in grade 5. It did not relate to either math or reading competency. Peer problems, however, were significantly related to math grades at year 5 and 7, as well as competency in grade 5 in both math and reading models.

Grades vs. Achievement Scores

As seen in **Figures 1, 2**, peer problems were predictive of grades broadly in both the reading and math models, and only of competency in the 5th grade. Meanwhile, prosocial behavior was significantly related to 5th year grades, but not 7th year, and never to competency.

We conclude that there is a greater overall relationship between grades and social behavior, particularly peer problems. Although, there is an indication of a relationship between peer problems and competency at an earlier grade.

TABLE 2 | Standardized factor loadings for peer problems and prosocial behavior.

SDQ items	Math M (SE)	Reading M (SE)
PEER PROBLEMS		
Item 3: Loner	0.40 (0.03)	0.38 (0.03)
Item 5: Has Friends	0.45 (0.03)	0.46 (0.03)
Item 6: Popular	0.48 (0.03)	0.46 (0.03)
Item 8: Is teased	0.61 (0.03)	0.59 (0.03)
Item 10: Gets along better with adults than with children	0.45 (0.03)	0.47 (0.03)
PROSOCIAL BEHAVIOR		
Item 1: Considerate	0.62 (0.02)	0.62 (0.02)
Item 2: Likes to share things	0.51 (0.02)	0.51 (0.03)
Item 4: Helpful	0.68 (0.02)	0.69 (0.02)
Item 7: Nice to younger children	0.51 (0.02)	0.51 (0.03)
Item 9: Often helps voluntarily	0.53 (0.02)	0.53 (0.02)

All loadings were significant at $p < 0.001$. All other loadings and path values for the math and reading models are visible in **Figures 1, 2**, respectively.

SES and Gender

Figures 1, 2 also indicate the effects of gender and parental education on competency and grades in both the math and reading models. Parental education was related to better grades and competency in both the math and reading models at both measurement points.

Gender was also a strong predictor of performance. Girls had worse math grades and competency than boys at both measurement points, and they had better grades than boys in both measurement points. However, they had better reading competency than boys in grade 7, but not at grade 5.

Overall, we conclude that gender and SES as determined by parental education correlated significantly with our dependent variables. Loadings from parental education appear to decrease from grades 5 and 7, and the effect of gender on reading became stronger between grades 5 and 7.

Social Behavior on Grades After the Controlling for Moderators

Both of the math and reading models modeled the variance attributed to gender and parental education separately from the variance of prosocial behavior and peer problems. A small to medium sized standardized path loading (path loadings between 0.06 and 0.12) on peer problems on grades and 5th grade competency remained. Thus, we can support our final prediction: that the relationship between social behavior and achievement remains despite including powerful moderating variables in our analyses.

DISCUSSION

Overview of Findings

Using data from a large-scale assessment of German students in early secondary schools, we provided evidence that social behavior has a disproportionate evidence on grades in comparison to achievement tests. Specifically these findings

help reconcile differential findings from studies using only grades or achievement tests as an outcome measure (e.g., (Adams et al., 1999; Malecki and Elliot, 2002; Lockl et al., 2017; Gerbino et al., 2018)). In our model, significant relationships between social behavior and both grades and early test scores, but not later test scores, remained. This remained true for both peer problems and prosocial behavior and true in both math and reading models.

Interpretation and Theoretical Implication

This novel finding was predicted by previous work which found noncognitive factors correlate more to grades than to IQ scores (Borghans et al., 2011; Lechner et al., 2017). The idea was further developed by Farrington et al. (2012), who identified social skills as one of several types of noncognitive factors influencing grades, one of which was social skills. Moreover, Farrington et al. (2012) called for future research to remedy to major insufficiencies in this line of research: research at the secondary level and research focusing on specific aspects of social skills. Our study addresses both these issues by examining early secondary students and by using the SDQ to define two specific dimensions of social skills: prosocial behavior and peer problems.

We further expand on the findings that internalizing problems are linked to reduced academic performance (Deighton et al., 2018) and that grades are also positively affected by prosocial behavior (Gerbino et al., 2018). One specific aspect of internalizing (i.e., peer problems) had a stronger negative impact on achievement, while prosocial behavior had a smaller positive effect only for grades. We also predicted a significant relationship between achievement test scores and social behavior, but were unable to support this prediction for math or reading beyond the 5th grade. Farrington et al. (2012) argued that social skills had an indirect effect and that it might be stronger for younger learners. Therefore, it is possible that the relationship between social behavior and competency fades as children age, or this relationship is too small to identify at later ages.

Limitations and Future Work

Despite our large and robust data set, some limitations remained. Our research focused on 5th and 7th graders. Full data from 9th grade and beyond in this cohort is not yet available. Thus, we cannot yet know the impacts of social behavior and skills on other life success measures and over a longer timeframe. One key assumption from Farrington et al. (2012) is that grades prove to be a better measure of future success, because they include noncognitive factors that are also important in long-term success. Therefore, future longitudinal research is necessary on this and similar cohorts to examine the hypothesis. Furthermore, given only two measurement points, it is difficult to make any causal inferences from this data. Broader longitudinal studies combined with intervention studies and true experiments are required to demonstrate a cause-and-effect connection.

Additionally, our research was further limited by only using limited aspects of social behavior. While prosocial behavior and peer problems are important, other aspects are also important for a full measure of social behavior, such as emotional competence, self-regulation, and aggression. While this data was not fully

available in this survey data, future research should endeavor to include additional specific measures of social behavior.

Another limitation comes from the types of data available in the NEPS database. While, the NEPS data-base includes self-reports of grades, it does not include self-reports of SDQ measures. Future work should compare the relationship between other sources of social behavior (e.g., self-report, parent report), and other sources of grades (e.g., teacher reports, academic records, etc.). Another artifact of the NEPS dataset is the order of the data collection. The SDQ subscales were collected between the achievement measures in our study, but we nonetheless treated them as predictors of both earlier and later achievement. Later studies may address this limitation by including more data from later measurement points, as those data become available.

Future work should work to integrate more variables into the analysis. We use a simplified rating of parental education to determine SES; however, parental education represents only a part of the SES, further work should incorporate other measures of SES such as income and living situation into analyses. Additional future work should also incorporate other personality variables, such as compliance, work ethic, and conscientiousness, which may have some overlap with our social behavior measures. Furthermore, the complex interaction of teacher expectation and support based on gender and SES and other variables should be considered. With the integration of these variables alongside an examination of the teacher-student interactions, the reasons for these effects could be further explained.

Lastly, although our dataset was broad and representative, it only included data from students attending schools in Germany. Future research is necessary on datasets from other nations as well as from multinational studies.

Application for Educational Practice

Our study further demonstrates the effect of social factors on grades and competency in math and reading. While there may be a potential bias effect on student grades for students based on prosocial behavior, this effect is small. Larger effects were observed for peer problems on both competency and grades. We recommend that teachers be aware of any social problems their students may possess as these learners may require additional support particularly in classrooms that use social learning styles.

REFERENCES

- Adams, J. W., Snowling, M. J., Hennessy, S. M., and Kind, P. (1999). Problems of behaviour, reading and arithmetic: assessments of comorbidity using the strengths and difficulties questionnaire. *Br. J. Educ. Psychol.* 69, 571–585. doi: 10.1348/000709999157905
- Andrei, F., Mancini, G., Mazzoni, E., Russo, P. M., and Baldaro, B. (2015). Social status and its link with personality dimensions, trait emotional intelligence, and scholastic achievement in children and early adolescents. *Learn. Individ. Differ.* 42, 97–105. doi: 10.1016/j.lindif.2015.07.014

CONCLUSION

Our goal was to examine the differential impact social behaviors (i.e., peer problems and prosocial behavior) on grades and achievement tests in both math and reading. Our results showed that grades correlate more strongly to social behavior than test scores do at younger ages, and that specifically peer problems have a stronger relationship to academic performance. Researchers should be careful when choosing which measure to use and especially when using both measures interchangeably. Teachers should likewise be aware of the relationship between social behavior and their students' grades. Future research into additional types of social behaviors and skills is necessary to identify the effects of specific aspects of social skills and behavior on specific grade types.

ETHICS STATEMENT

This study used existing data from the German National Education Panel Study. From 2008 to 2013, data collection was supervised by the Framework Programme for the Promotion of Empirical Education Research. As of 2014, data collection was carried out by the Leibniz Institute for Educational Trajectories. Because the study used existing data, no new ethical review was required.

AUTHOR CONTRIBUTIONS

JD served as primary author and data analyst. MG provided writing oversight, feedback, and initial study design, KR provided expertise on developing, modeling, and utilizing control variables (i.e., gender, SES) and theoretical expertise in their implementation in NEPS and their theoretical discussion.

ACKNOWLEDGMENTS

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 3–5th Grade, doi: 10.5157/NEPS:SC3:6.0.1. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network.

- Askell-Williams, H., and Lawson, M. J. (2015). Relationships between students' Mental health and their perspectives of life at school. *Health Educ.* 115, 249–268. doi: 10.1108/HE-02-2014-0007
- Baroody, A. E., Rimm-Kaufman, S. E., Larsen, R. A., and Curby, T. W. (2016). A multi-method approach for describing the contributions of student engagement on fifth grade students' social competence and achievement in mathematics. *Learn. Individ. Differ.* 48, 54–60. doi: 10.1016/j.lindif.2016.02.012
- Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (eds.). (2011). *Education as a Lifelong Process – The German National Educational Panel Study (NEPS). [Special Issue] Zeitschrift für Erziehungswissenschaft*, 14.

- Borghans, L., Golsteyn, B. H., Heckman, J., and Humphries, J. E. (2011). Identification problems in personality psychology. *Pers. Individ. Differ.* 51, 315–320. doi: 10.1016/j.paid.2011.03.029
- Bradley, R. H., and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annu. Rev. Psychol.* 53, 371–399. doi: 10.1146/annurev.psych.53.100901.135233
- Brauns, H., Scherer, S., and Steinmann, S. (2003). “The CASMIN educational classification in international comparative research,” in *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, eds J. H. P. Hoffmeyer-Zlotnik and C. Wolf (Boston, MA: Springer), 221–244.
- Card, N. A., Stucky, B. D., Sawalani, G. M., and Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: a meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Dev.* 79, 1185–1229. doi: 10.1111/j.1467-8624.2008.01184.x
- Chaplin, T. M., and Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* 139, 735–765. doi: 10.1037/a0030737
- Cross, L. H., and Frary, R. B. (1999). Hodgepodge grading: endorsed by students and teachers alike. *Appl. Measure. Educ.* 12, 53–72.
- Deighton, J., Humphrey, N., Belsky, J., Boehnke, J., Vostanis, P., and Patalay, P. (2018). Longitudinal pathways between mental health difficulties and academic performance during middle childhood and early adolescence. *Br. J. Dev. Psychol.* 36, 110–126. doi: 10.1111/bjdp.12218
- DeVries, J. M., Gebhardt, M., and Voß, S. (2017). An assessment of measurement invariance in the 3- and 5-factor models of the strengths and difficulties questionnaire: new insights from a longitudinal study. *Pers. Individ. Differ.* 119, 1–6. doi: 10.1016/j.paid.2017.06.026
- Durlak, J. A., Weissberg, R. P., and Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *Am. J. Commun. Psychol.* 45, 294–309. doi: 10.1007/s10464-010-9300-6
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching Adolescents to Become Learners. The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review*. Chicago: University of Chicago Consortium on Chicago School Research.
- Gerbino, M., Zuffianò, A., Eisenberg, N., Castellani, V., Luengo Kanacri, B. P., Pastorelli, C., et al. (2018). Adolescents’ prosocial behavior predicts good grades beyond intelligence and personality traits. *J. Pers.* 86, 247–260. doi: 10.1111/jopy.12309
- Goodman, A., Lamping, D. L., and Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *J. Abnorm. Child Psychol.* 38, 1179–1191. doi: 10.1007/s10802-010-9434-x
- Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *J. Child Psychol. Psychiatry* 38, 581–586. doi: 10.1111/j.1469-7610.1997.tb01545.x
- Haberkorn, K., Pohl, S., Hardt, K., and Wiegand, E. (2012). *NEPS Technical Report for Reading – Scaling Results of Starting Cohort 3 in Fifth Grade* (NEPS Working Papers No. 15). Bamberg. Available online at: https://www.neps-data.de/Portals/0/Working%20Papers/WP_XV.pdf
- Hu, L.-T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424
- Krannich, M., Jost, O., Rohm, T., Koller, I., Carstensen, C. H., Fischer, L., et al. (2017). *NEPS Technical Report for Reading - Scaling Results of Starting Cohort 3 for Grade 7* (NEPS SURVEY PAPERS No. 14). Bamberg.
- Lechner, C., Danner, D., and Rammstedt, B. (2017). How is personality related to intelligence and achievement? A replication and extension of Borghans et al. and Salkever. *Pers. Individ. Differ.* 111, 86–91. doi: 10.1016/j.paid.2017.01.040
- Lekholm, A. K., and Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educ. Res. Eval.* 14, 181–199. doi: 10.1080/13803610801956663
- Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., and Young-Morris, C. (2013). Socioeconomic status and child development: a meta-analysis. *J. Emot. Behav. Disord.* 21, 211–224. doi: 10.1177/1063426611421007
- Lewis, G. J., Asbury, K., and Plomin, R. (2017). Externalizing problems in childhood and adolescence predict subsequent educational achievement but for different genetic and environmental reasons. *J. Child Psychol. Psychiatry* 58, 292–304. doi: 10.1111/jcpp.12655
- Lockl, K., Ebert, S., and Weinert, S. (2017). Predicting school achievement from early theory of mind: differential effects on achievement tests and teacher ratings. *Learn. Individ. Differ.* 53, 93–102. doi: 10.1016/j.lindif.2016.11.007
- Malecki, C. K., and Elliot, S. N. (2002). Children’s social behaviors as predictors of academic achievement: a longitudinal analysis. *School Psychol. Q.* 17, 1–23. doi: 10.1521/scpq.17.1.1.19902
- Muthén, L. K., and Muthén, B. O. (1998–2017). *Mplus User’s Guide, 8th Edn*. Los Angeles, CA: Muthén & Muthén.
- Piff, P. K., and Robinson, A. R. (2017). Social class and prosocial behavior: current evidence, caveats, and questions. *Curr. Opin. Psychol.* 18, 6–10. doi: 10.1016/j.copsyc.2017.06.003
- Robinson, J. P., and Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: examining direct cognitive assessments and teacher ratings. *Am. Educ. Res. J.* 48, 268–302. doi: 10.3102/0002831210372249
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417
- Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice, 2nd Edn*. Boston, MA: Allyn and Bacon.
- Slavin, R. E. (2014). Cooperative learning in elementary schools. *Education* 43, 5–14. doi: 10.1080/03004279.2015.963370
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., and Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin. Child Fam. Psychol. Rev.* 13, 254–274. doi: 10.1007/s10567-010-0071-2
- Veland, J., Bru, E., and Idsøe, T. (2015). Perceived socio-economic status and social inclusion in school: parental monitoring and support as mediators. *Emot. Behav. Difficulties* 20, 173–188. doi: 10.1080/13632752.2014.931018
- Vygotsky, L. (1978). “Interaction between learning and development,” in *Readings on the Development of Children*, eds M. Gauvain and G. Cole (New York, NY: Scientific American Books), 34–40.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *J. Educ. Psychol.* 85, 357–364. doi: 10.1037/0022-0663.85.2.357
- Wentzel, K. R. (1998). Social relationships and motivation in middle school: the role of parents, teachers, and peers. *J. Educ. Psychol.* 90, 202–209. doi: 10.1037/0022-0663.90.2.202
- Wentzel, K. R., and Caldwell, K. (1997). Friendships, peer acceptance, and group membership: relations to academic achievement in middle school. *Child Dev.* 68, 1198–1209. doi: 10.1111/j.1467-8624.1997.tb01994.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 DeVries, Rathmann and Gebhardt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Mplus instructions for the reading model. The math model was identical, except for substituting math for reading.

Variable:

Names are tw2_1-tw2_10, gender, rg5, rg7, ParEd, rc5, rc7,
SchoolID;
Usevariables are tw2_1-tw2_10, gender, rg5 rg7 ParEd rc5
rc7;
missing are all (-99 - -2);
Cluster is SchoolID;

Model:

!SDQ subscales
PP2 by tw2_3 tw2_5 tw2_6 tw2_8 tw2_10;
PrS2 by tw2_1 tw2_2 tw2_4 tw2_7 tw2_9;

!Achievement measures on social factors
rg5 rc5 rg7 rc7 on PP2;
rg5 rc5 rg7 rc7 on PrS2;
!Control Variables
PrS2 PP2 rc5 rc7 rg5 rg7 on Gender;
PrS2 PP2 rc5 rc7 rg5 rg7 on ParEd;
Gender with ParEd@0;
!Achievement measures - competency predicting grades
rg5 on rc5;
rg7 on rc7;
!Achievement measures - Grade 5 to Grade 7 regression
rg7 on rg5;
rc7 on rc5;

Analysis:

type is complex;
estimator is MLR;

Output:

stdyx;
sampstat;

Analysis:

type is complex;
estimator is MLR;



Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Do learners with special education needs really feel included? Evidence from the Perception of Inclusion Questionnaire and Strengths and Difficulties Questionnaire

Jeffrey M. DeVries^{a,*}, Stefan Voß^b, Markus Gebhardt^a^a Technical University of Dortmund, Emil-Figge-Str. 50, 44227 Dortmund, Germany^b University of Rostock, Albert-Einstein-Str. 21, 18059 Rostock, Germany

ARTICLE INFO

Keywords:

Special education needs
 Inclusive education
 Strengths and Difficulties Questionnaire
 Perception of Inclusion Questionnaire
 Academic self-concept
 Self-perception

ABSTRACT

Background: School inclusion is an important right of students in school systems around the world. However, many students with special education needs (SEN) have lower perceptions of inclusion despite attending inclusive schools.

Aims: This study examined perceived levels of inclusion, academic self-concept and developmental problems in inclusive schools.

Methods and procedures: The Strengths and Difficulties Questionnaire and the Perception of Inclusion Questionnaire were administered at two measurement points (6th and 7th grade; $n = 407$, including 48 with SEN) at multiple inclusive schools. Responses were compared based on gender, grade level, and SEN. Factor structure and measurement invariance were evaluated. **Outcomes and results:** Factor structures of both questionnaires were confirmed. Academic self-concept and emotional inclusion were lower for learners with SEN. However, these effects shrank in grade 7. Similarly, academic self-concept increased between grade 6 and 7. Lastly, learners with SEN had a higher level of conduct problems.

Conclusions and implications: Both instruments remain suitable for use in comparisons in inclusive schools. Significant differences exist for learners with SEN in inclusive classrooms, although these differences may shrink over time. We recommend the continued use of the Perception of Inclusion Questionnaire for information about school inclusion and for learners with SEN.

What this paper adds?

This paper contributes to the literature on inclusion in three important ways. First, the study examines participants across multiple schools in an inclusive school system, providing an *in situ* measurement of how included learners both with and without special education needs (SEN) perceive themselves to be included in their classes. Second, this paper uses an important new instrument to assess perception of inclusion, the perception of inclusion questionnaire (PIQ). The PIQ is evaluated alongside a well-researched instrument, the strengths and difficulties questionnaire (SDQ), allowing for a comparison between both scales. Besides data on the social inclusion and the emotional inclusion, the PIQ provides valuable information about the self-concept of learners, which the SDQ does not assess. Further, we examine the factor structure and invariance of both measures across SEN, gender, and grade level. Third, we found that learners with SEN feel a lower academic self-concept and feel less emotionally included across both

* Corresponding author.

E-mail address: jeffrey.devries@tu-dortmund.de (J.M. DeVries).

<https://doi.org/10.1016/j.ridd.2018.07.007>

Received 8 March 2018; Received in revised form 17 July 2018; Accepted 19 July 2018
 0891-4222/ © 2018 Elsevier Ltd. All rights reserved.

measurement points and less socially included in the 6th grade (although not the 7th grade). Further, the difference between these groups shrinks from 6th to 7th grade. This novel interaction was unaccounted for in previous research; therefore, it necessitates more work to investigate the nature of inclusive schooling related to perceived social and emotional inclusion for children with SEN.

1. Introduction

While inclusion in schools is an agreed-upon international goal (United Nations Convention on the Rights of Persons with Disabilities, 2006), many students with special education needs (SEN) remain excluded socially and emotionally from regular classroom experiences (Banks, McCoy, & Frawley, 2017; Bossaert, Colpin, Pijl, & Petry, 2013; Prince & Hadwin, 2013; Schwab, Gebhardt, & Gasteiger-Klicpera, 2013). Similarly, learners with SEN are at risk for a lower academic self-concept (Elbaum, 2002; Li, Tam, & Man, 2006; Wei & Marder, 2010). Emotional, social, and academic inclusion at school can reduce the negative risks faced by such students with SEN (Grütter, Gasser, & Malti, 2017; Schwab, 2017). However, differing inclusive schools may vary on how much access learners with SEN may have and how much support such students receive, resulting in greater or lower benefits from inclusion (Prince & Hadwin, 2013).

This study will examine students in inclusive schools in order to identify the relationship between higher perceived inclusion and emotional, social, and conduct problems. We will use a well-established instrument, the strengths and difficulties questionnaire (SDQ; Goodman, 1997) alongside a newer instrument, the perceptions of inclusion questionnaire (PIQ; Venetz, Zurbriggen, & Eckhart, 2014). We will evaluate these instruments' applicability for learners with and without SEN and we will examine how both instruments relate to gender and SEN across two measurement points (6th and 7th Grade).

1.1. Socioemotional inclusion and academic self-concept

School inclusion is related to social, affective, and self-concept outcomes for children with SEN (for a review, see Prince & Hadwin, 2013). Learners with SEN may lack key social skills (Schwab et al., 2013; Wight & Chapparo, 2008). They may face greater prejudice (Avramidis, 2010) and bullying (Rose, Monda-Amaya, & Espelage, 2010), and are at a greater risk to feel excluded at schools (Grütter et al., 2017; Schwab et al., 2013; Schwab, 2017). Recent large-scale studies found that students with SEN felt less included at school (Stiefel, Shiferaw, Schwartz, & Gottfried, 2017) and that they had fewer friends and experienced negative peer relationships than other students (Avramidis, Avgeri, & Strogilos, 2018; Banks et al., 2017; Huber, Gerullis, Gebhardt, & Schwab, 2018). A lack of inclusion is also related to many negative emotional-developmental outcomes, including depression, (McGraw, Moore, Fuller, & Bates, 2008), substance abuse, and other mental health problems (Arslan, 2018; Bond et al., 2007), as well as poorer academic outcomes (Szumski & Karwowski, 2015; Van Ryzin, Gravely, & Roseth, 2009) such as a reduced academic self-concept (Bear, Minke, & Manning, 2002). Children with SEN are at an even greater risk of poor academic outcomes due to worse feelings of inclusion and a lower academic self-concept (Elbaum & Vaughn, 2003; Korhonen, Linnanmäki, & Aunio, 2014; Szumski & Karwowski, 2015).

1.2. The Perceptions of Inclusion Questionnaire

The Perceptions of Inclusion Questionnaire (PIQ; Venetz et al., 2014; Venetz, Zurbriggen, Eckhart, Schwab, & Hessels, 2015) is designed to measure three areas critical to inclusion: social inclusion, emotional inclusion, and academic self-concept. Academic self-concept (see Elbaum, 2002; Elbaum & Vaughn, 2003; Prince & Hadwin, 2013) describes a learner's self-concept in the specific domain of school. Relatedly, emotional inclusion refers to a sense of well-being at school and social inclusion describes the sense of connectedness (e.g., friends) at school (Elbaum & Vaughn, 2003; Schwab et al., 2013; Stiefel et al., 2017; Szumski & Karwowski, 2015). The PIQ measures these three constructs (academic self-concept, emotional inclusion, and social inclusion) with four items each on a 4-point Likert scale ("not at all true" to "certainly true"). It can be self-administered or taken by a child's teacher or parent, and it is designed for 8- to 16-year-olds.

The PIQ is based on the longer questionnaire to assess the dimensions of integration of pupils (FDI; in German: Fragebogen zur Erfassung der Dimensionen der Integration von Schülern; Haeblerlin, Moser, Bless, & Klaghofer, 1989). The FDI was the first instrument developed in Switzerland to assess levels of perceived inclusion by students at schools. It had a big influence in the discussion of implementation of inclusion in the German speaking countries (Gebhardt, Schwab, Krammer, & Gasteiger, 2012; Sauer, Ide, & Borchert, 2007; Schwab et al., 2013; Schwab, Gebhardt, Krammer, & Gasteiger-Klicpera, 2015). The FDI included 45 items across 3 factors, and the PIQ was able to reduce the scale to 12 items across the same 3 factors (4 items per factor), while maintaining a high Cronbach's alpha (all $\alpha \geq .80$) and good model fits (Venetz et al., 2014).

English, German, and other language versions of the PIQ are available online to educators and researchers (see Venetz et al., 2015). The scale was further evaluated by Zurbriggen, Venetz, Schwab, and Hessels (2017), where its 3-factor structure was confirmed. Overall, the test items behaved normally. However, one item was found to lack measurement invariance between learners with learning disabilities and those without. Zurbriggen et al. (2017) concluded that more work is necessary to compare the instrument in general across differing SEN. Furthermore, a comparison of the scale with other established scales will allow for additional cross-validation of its latent variables. This study seeks to close this research gap by comparing the PIQ with the SDQ.

1.3. The Strengths and Difficulties Questionnaire

The SDQ is a well-established tool for assessing individual personality qualities of 2- to 17-year old children (Goodman, 1997; Goodman, Lamping, & Ploubidis, 2010). It is traditionally assessed via a 5-factor structure matching to its five subscales: emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behavior (Goodman et al., 2010). Each subscale contains five items with three response categories describing the frequency of a behavior (0: not true, 1: somewhat true, 2: certainly true). It possesses measurement invariance over time in a sample of German children (DeVries, Gebhardt, & Voß, 2017); however, its invariance across SEN and gender in the same sample have not been demonstrated. Because the SDQ is used as a clinical screening test (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000; Mellins et al., 2018), an evaluation of its invariance across SEN is necessary.

1.4. Present study

As described by Prince and Hadwin (2013), low levels of inclusion relate to increased risk for social and emotional problems, especially for children with SEN. The SDQ will allow us to examine socioemotional problems alongside perceptions of emotional and social inclusion as well as academic self-concept from the PIQ. This comparison will enable us to connect the subscales of the PIQ to established theory. Furthermore, we will examine the functioning of the PIQ and SDQ across different subject groups (i.e., gender, SEN, and grade). In the present study, we compare the sum (SDQ) and mean scores (PIQ) on these scales; we assess their metric invariance over grades 6 and 7, over the presence of SEN, and over gender; and we examine the effects of grade level, SEN, and gender on both instruments. These research goals led us to our five research questions.

- 1 Do the subscales of the PIQ and SDQ correlate to each other?
- 2 Do the SDQ and PIQ possess measurement invariance across 6th and 7th grade, SEN, and gender?
- 3 Do students' answers differ on the PIQ and SDQ between 6th and 7th grade?
- 4 Do boys and girls respond differently to the PIQ and SDQ?
- 5 Do learners with SEN respond differently to PIQ and SDQ?

2. Methods

2.1. Participants

Participants were 407 (47.7% girls) students attending 9 different inclusive middle schools in a school district in northern Germany. Each school had between 1 and 3 classes with 14–27 students per class. The sample included 48 learners with SEN. Special education needs were diagnosed by the responsible diagnostic service based on the ICD-10 classification from the World Health Organization (WHO), category F81 “Specific developmental disorders of scholastic skills” ($n = 29$), learners who repeated a grade ($n = 17$), and students who were referred for diagnosis by their teachers, but were not officially diagnosed ($n = 2$). The students with SEN attended mainstream schools. In this school district an agreement between these schools and government agencies established a framework for the inclusion of learners with SEN (Voß et al., 2017). These include:

- Flexible support for students with minor academic problems by their teachers (e.g., extended learning time through additional small groups instruction)
- Additional support for students with serious persistent problems by special educators (e.g., individualized education plan, specific trainings)
- A twice-yearly evaluation of the academic achievement of all students via standardized tests in language and mathematics in order to foster instruction, decision-making, and school resource utilization
- At least half-yearly team consultations between subject teachers and special education teachers about the status of learners with SEN
- Adverse balancing or goal-differentiated evaluation and censoring for students with SEN

According to this agreement, about 0.3 teacher lessons per student are available for special educational support. This corresponds to about 6–8 h per week in an average class.

2.2. Instruments & procedure

Students responded on the self-report versions of the German language SDQ and PIQ. Self-reports were chosen over teacher ratings because secondary-school students would have multiple teachers throughout the day and over two years. Self-reports are also easier to administer than parent ratings through the process of regular testing. Teachers administered the questionnaires in the regular classroom, to all students simultaneously.

Ratings were collected at the end of the 6th and 7th grade. However, only a reduced sample was available for the 6th grade ($n = 288$) due to a coding error. A series of t -tests comparing 7th grade SDQ and PIQ subscales confirmed no systematic differences existed between those included ($n = 288$) and excluded ($n = 119$) in the 6th grade (all $ps > .05$). For within-subjects analyses, we used the reduced dataset of 288, but for all between-subjects analyses, we used the full sample of 407 available at the second

Table 1
Factor Structure & Standardized Factor Loadings of the Perception of Inclusion Questionnaire (PIQ).

Item	Factor & Full Item Text	Loading
	Emotional Inclusion	
1	I like going to school.	0.875
4	<i>I have no desire to go to school.</i>	0.733
7	I like it in school.	0.782
10	School is fun.	0.861
	Social Inclusion	
2	I have a lot of friends in my class.	0.677
5	I get along very well with my classmates.	0.740
8	<i>I feel alone in my class.</i>	0.618
11	I have very good relationships with my classmates.	0.772
	Academic Self-Concept	
3	I am a fast learner.	0.765
6	I am able to solve very difficult exercises.	0.707
9	I do well in my schoolwork.	0.761
12	<i>Many things in school are too difficult for me.</i>	0.588

Note: Italicized items are reverse scored. All factor loadings were significant at $p < .001$. Full version of the PIQ is available under Venetz et al. (2015).

measurement point.

2.2.1. The Perceptions of Inclusion Questionnaire (PIQ)

The PIQ (see Venetz et al., 2015 for full version) is a short-form version of the questionnaire to assess the dimensions of integration of pupils (FDI; originally: Fragebogen zur Erfassung von Dimensionen der Integration von Schülern FDI 4–6; Haeberlin et al., 1989). It includes the three scales: emotional integration, social integration, and academic self-concept. Each scale is significantly related to the inclusion of a child in the school system (Venetz et al., 2014; Zurbriggen et al., 2017). The PIQ items and factors are described in Table 1.

2.2.2. The Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a 25-item questionnaire that rates children on conduct problems, peer problems, hyperactivity, emotional problems, and prosocial behavior. It has been used in many countries (Bøe, Hysing, Skogen, & Breivik, 2016; Ortuño-Sierra, Fonseca-Pedrero et al., 2015) and across many age groups (e.g., DeVries et al., 2017; Hagquist, 2007). The SDQ items and factors are described in Table 2.

2.3. Analyses

2.3.1. Factor analyses

Separate confirmatory factor analyses (CFAs) for the PIQ and SDQ were conducted in Mplus 7.4 (Muthén & Muthén, 1998–2015) using robust maximum likelihood estimation (MLR). Fits for both instruments at the second measurement point were assessed. Root mean square of approximation (RMSEA) < 0.08 , comparative fit index (CFI) $> .90$, and standardized root mean square residual (SRMR) $< .10$ were considered acceptable fits, and RMSEA $< .05$, CFI $> .95$, and SRMR $< .08$ were considered good fit values (Brown, 2015; Hu & Bentler, 1999). Minor model modifications for the SDQ to reach acceptable fits are detailed in the results section. Similar minor modifications for SDQ data are common (e.g., Ortuño-Sierra, Fonseca-Pedrero et al., 2015).

Next, separate invariance measures were calculated for both instruments in Mplus. We tested invariance across measurement points, gender, and presence of SEN. For the gender and SEN invariance tests, we used data from the 7th grade because of the larger number of participants. Weak and strong invariance were assessed by comparing the base, metric and scalar models following the procedures recommended by Dimitrov (2017). Changes in MLR corrected χ^2 (Satorra & Bentler, 2010) resulting in a significant difference ($p < .05$) were set as our critical threshold. In the event of invariance, we examined the effects of constraining specific factor loadings and freeing specific intercepts. If only 20% of loadings and intercepts were not invariant we concluded we had partial invariance sufficient for subsequent comparisons (see Levine et al., 2003).

2.3.2. Correlations

In order to examine the relationship between the two scales, sum scores (SDQ) and mean scores (PIQ) were compared in a correlation matrix for measurement point 2.

2.3.3. MANOVA

Lastly, we examined the effects of gender and SEN over both measurement points in a repeated measures MANOVA (Gender x SEN x Grade) with each subscale as dependent variable.

Table 2
Factor Structure & Standardized Factor Loadings of the Strengths and Difficulties Questionnaire (SDQ).

Item	Factor & Short Item Text	Loading
	Emotional Problems	
3	I get a lot of headaches...	0.449
8	I worry a lot	0.563
13	I am often unhappy...	0.646
16	I am nervous in new situations...	0.503
24	I have many fears...	0.610
	Conduct Problems	
5	I get very angry	0.549
7	<i>I usually do as I am told</i>	0.427
12	I fight a lot	0.480
18	I am often accused of lying or cheating	0.351
22	I take things that are not mine	0.443
	Hyperactivity	
2	I am restless...	0.732
10	I am constantly fidgeting...	0.663
15	I am easily distracted	0.586
21	<i>I think before I do things</i>	0.417
25	<i>I finish the work I am doing</i>	0.377
	Peer Problems	
6	I am usually on my own	0.539
11	<i>I have one good friend or more</i>	0.543
14	<i>Other people my age generally like me</i>	0.396
19	Other children or young people pick on me	0.663
23	I get along better with adults than with people my age	0.409
	Prosocial Behavior	
1	I try to be nice to other people	0.630
4	I usually share with others	0.376
9	I am helpful if someone is hurt...	0.638
17	I am kind to younger children	0.509
20	I often volunteer to help others	0.490

Note: Italicized items were reverse scored. In the modified model, item 18 was cross-loaded onto peer problems with a standardized loading of 0.341 and item 8 was cross-loaded onto prosocial behavior with a factor loading of 0.287. All loadings, including modifications, were significant at $p < .001$.

3. Results

3.1. Model fits & model modifications

The 3-factor structure of the PIQ had good fit metrics, RMSEA = .053 (0.42–.064), CFI = .952, and SRMR = 0.058. While its RMSEA was above the 0.05 threshold for good fits, this value was in the acceptable range, while other values were in the good ranges. Furthermore, as seen in Table 1, all factor loadings were significant.

Initial 5-factor SDQ fits were insufficient, RMSEA = .043 (.036–.049), CFI = .866, SRMR = .064. While the RMSEA and SRMR were in the good range, the CFI was below our threshold for acceptable fit. Therefore, three minor modifications were made based on reported modification indices and theoretically related items. Specifically, item 8 was cross listed under prosocial behavior, item 18 was cross listed under peer problems, and the errors of items 20 and 25 were correlated. In the self-report version, item 8 reads “I worry a lot,” which might relate to a greater motivation to help others. Similarly, item 18 reads “I am often accused of lying or cheating.” This may relate to peer-problems because accusations are also affected by social relationships. Lastly, item 20, “I often volunteer to help,” and item 25, “I Finish the work I am doing,” share aspects of conscientiousness (see [Ortuño-Sierra, Chocarro, Fonseca-Pedrero, Riba, & Muñiz, 2015](#) for detailed description of similar modifications to the SDQ). The modified model had sufficient fit metrics, with RMSEA = .035 (.027–.042), CFI = .912, and SRMR = .056. Additionally, as seen in Table 2, all factor loadings were significant.

3.1.1. Invariance tests

Table 3 describes the results of the invariance analysis. Both the SDQ and PIQ possessed strong invariance across both grade level and SEN. Furthermore, the PIQ showed strong invariance across gender, but the SDQ lacked invariance across gender.

Further examination of factor loadings and intercepts revealed that items 3, 5, 9, 13, and 22 of the SDQ resulted in significantly worse fits with their loadings freed, and items 1, 2, 9, 11, 18, 20, 22, and 22 resulted in significantly better fits with their intercepts freed. Thus, 13 of 50 (26%) of loadings and intercepts were not invariant. Because the established threshold for partial invariance is 20% ([Dimitrov, 2017](#); [Levine et al., 2003](#)), we cannot reasonably conclude sufficient partial invariance exists in the SDQ across gender.

Table 3
Invariance Tests for the Strengths and Difficulties Questionnaire and Perception of Inclusion Questionnaire.

	Base Model χ^2 (df)	Metric Model χ^2 (df)	Scalar Model χ^2 (df)	Metric vs. Base Model p-value	Scalar vs. Metric Model p-value
Grade Level					
SDQ	814.96 (524)	832.57 (546)	862.05 (566)	.588	.081
PIQ	189.62 (102)	201.41 (111)	217.92 (120)	.246	.057
SEN					
SDQ	760.08 (524)	777.68 (546)	801.61 (566)	.400	.261
PIQ	218.49 (102)	233.31 (111)	247.49 (120)	.093	.150
Gender					
SDQ	775.58 (524)	812.09 (546)	865.62 (566)	.035*	.001**
PIQ	197.63 (102)	205.71 (111)	218.93 (120)	.524	.177

Note: Base model has free loadings and intercepts between groups. Metric model has equal loadings, but free intercepts, scalar model has equal loadings and intercepts. All SDQ comparisons use the model with minor modifications.

* Significant at $p < .05$.

** Significant at $p < .01$.

3.2. Correlations

Table 4 shows a correlation matrix between each of the subscales for both instruments. All PIQ subscales of academic inclusion were positively correlated with prosocial behavior and negatively correlated to difficulties items from the SDQ. All correlations were significant except for social inclusion with hyperactivity and emotional inclusion with peer problems.

3.3. Grade level, special education needs, and gender

3.3.1. Grade level

Marginal means from the MANOVA can be found in Table 5. A significant main effect for grade level was only found for academic self-concept, $F(1, 212) = 13.55, p < .001$, partial $\eta^2 = .060$. This medium-sized effect is further explained by the interaction between special education needs and grade level (see below). There were no other main effects of grade level ($p > .05$).

3.3.2. Special education needs

As seen in Table 5, children with SEN had a significantly lower academic self-concept, $F(1, 212) = 17.13, p < .001$, partial $\eta^2 = .075$. Additionally, they had a lower feeling of emotional inclusion, $F(1,212) = 4.143, p < .05$, partial $\eta^2 = .019$. Similarly, they had a greater incidence of conduct problems, $F(1, 212) = 5.083, p < .05$, partial $\eta^2 = .023$. This indicates a medium effect of SEN on academic self-concept, and small effects of SEN on emotional inclusion and conduct problems. There were no other main effects of SEN ($p > .05$).

3.3.3. Gender

Also seen in Table 5, there were no significant effects of gender on the PIQ subscales ($p > .05$). Although female students did have a significantly greater level of prosocial behavior and emotional symptoms, comparisons by gender on the SDQ are not reliable because we could not establish measurement invariance across gender for the instrument.

Table 4
Correlation Matrix of subscales of the Strengths and Difficulties Questionnaire and Perception of Inclusion Questionnaire.

	ASC	SI	EI	ES	CP	HA	PP	PB
PIQ Academic Self-Concept	1.000							
PIQ Social Inclusion	.184***	1.000						
PIQ Emotional Inclusion	.360***	.231***	1.000					
SDQ Emotional Symptoms	-.236***	-.202***	-.109*	1.000				
SDQ Conduct Problems	-.243***	-.244***	-.341***	.192***	1.000			
SDQ Hyperactivity	-.389***	.080	-.441***	.144**	.504***	1.000		
SDQ Peer Problems	-.212***	-.619***	-.085	.373***	.305***	.080	1.000	
SDQ Prosocial Behavior	.267***	.309***	.449***	.028	-.340***	-.324***	-.232***	1.000

Note: Correlations for all values at grade 7.

* Significant at $p < .05$.

** Significant at $p < .01$.

*** Significant at $p < .001$.

Table 5
MANOVA – Marginal Means & Main Effects.

	Grade 6 M (SE)	Grade 7 M (SE)	p-value	Male M (SE)	Female M (SE)	p-value	No SEN M (SE)	With SEN M (SE)	p-value
PIQ									
Academic Self-Concept	2.50 (0.07)	2.72 (0.07)	.001***	2.72 (0.08)	2.50 (0.10)	.073	2.86 (0.05)	2.36 (0.11)	.001***
Social Inclusion	3.27 (0.07)	3.34 (0.06)	.701	3.26 (0.07)	3.35 (0.08)	.403	3.32 (0.04)	3.29 (0.10)	.779
Emotional Inclusion	2.47 (0.08)	2.50 (0.08)	.288	2.42 (0.09)	2.55 (0.12)	.372	2.64 (0.05)	2.34 (0.14)	.043*
SDQ									
Emotional Symptoms	2.94 (0.20)	3.35 (0.23)	.058	2.50 (0.23)	3.78 (0.30)	.001**	3.03 (0.14)	3.26 (0.35)	.527
Conduct Problems	2.16 (0.16)	1.95 (0.15)	.229	2.28 (0.16)	1.83 (0.21)	.088	1.76 (0.10)	2.36 (0.25)	.025*
Hyperactivity	4.18 (0.23)	4.17 (0.22)	.948	4.51 (0.24)	3.84 (0.30)	.081	3.96 (0.14)	4.39 (0.36)	.273
Peer Problems	2.76 (0.21)	2.53 (0.18)	.284	2.77 (0.20)	2.53 (0.28)	.467	2.54 (0.12)	2.75 (0.30)	.516
Prosocial Behavior	7.17 (0.18)	7.10 (0.19)	.689	6.70 (0.20)	7.58 (0.25)	.006**	7.34 (0.12)	6.94 (0.30)	.214

Note: SEN refers to special education needs. Significant main effects for gender on the SDQ subscales are not trustworthy due to a lack of measurement invariance.

*** Significant at $p < .001$.

** Significant at $p < .01$.

* Significant at $p < .05$.

3.3.4. Interactions

Significant time by SEN interactions were found for comparisons on each subscale of the PIQ, academic self-concept $F(1, 212) = 6.45, p < .05$, partial $\eta^2 = .030$; emotional inclusion $F(1,212) = 4.60, p < .05$, partial $\eta^2 = .021$; and social inclusion $F(1, 212) = 4.69, p < .05$, partial $\eta^2 = .022$. As seen in Fig. 1, the academic self-concept, emotional inclusion, and social inclusion of children with SEN increases from grade 6 to 7. Meanwhile, these values remain flat (academic self-concept) or decrease (emotional inclusion and social inclusion) for children without SEN. No other interactions were detected in our MANOVA ($p > .05$).

4. Discussion

We examined the relationship between self-reported socioemotional problems and perceived school inclusion in a sample taken from inclusive schools. We further examined the effects of grade level, special education needs, and gender on socioemotional difficulties and perception of inclusion. Correlations between SDQ and PIQ scales indicated that children with higher SDQ scores for peer, emotional, and conduct problems perceive themselves to be less included in the classroom. This was also true for children with SEN (as described in Section 2.1), which was demonstrated by lower levels of academic self-concept and emotional inclusion on the PIQ.

However, in our sample from inclusive schools, learners with SEN improve on perceptions of academic self-concept, social inclusion, and emotional inclusion from grade 6 to grade 7. By Grade 7, social inclusion of both groups is indistinguishable, and the differences in emotional inclusion shrinks dramatically (i.e., overlapping error bars). Learners with SEN at grade 7 still have a lower academic self-concept, but the difference has significantly shrank. This finding is not predicted by previous research on social inclusion which found a lower overall level of social inclusion in learners with SEN (Bossaert et al., 2013; Schwab et al., 2013; Stiefel et al., 2017). However, this may be related to the longitudinal effect of inclusive schooling, which may boost the academic self-concept, as well as social and emotional well-being of learners with SEN. Alternatively, this could be an effect of the school transition

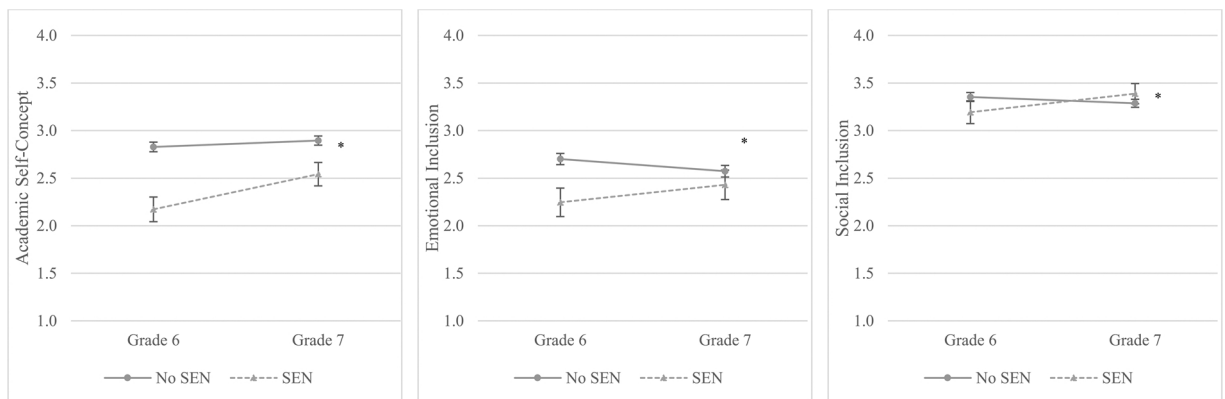


Fig. 1. Grade Level x Special Education Needs Interactions from the Perception of Inclusion Questionnaire.

Note: SEN refers to Special Education Needs.

* Significant at $p < .05$.

from elementary to middle school levels. Students may have had an artificially lowered self-concept at grade 6 (the year after transition). Another possibility is that of actual improved academic performance, reflected by improved self-concept. More longitudinal research is necessary to determine the specific cause of this interaction. Such research should assess changes across different school types, transitions, and make comparisons to school achievement.

It is important to note that this is unlikely to be an artifact of the PIQ. The subscales of the PIQ correlate to the SDQ's subscales in the expected directions. In other words, children with more socioemotional problems had lower academic self-concept and lower feelings of inclusion. This is congruent with predictions based on Prince and Hadwin's (2013) synthesis, which concluded that socioemotional problems correlate to lower levels of inclusion at school. This is of greater importance for children with SEN, who face lower levels of inclusion already. Furthermore, despite the relationship between both instruments, the PIQ demonstrated significant differences for learners with SEN, while the SDQ found these differences only for conduct problems. One possibility is that the PIQ is more sensitive to disruptions caused by special education needs. Furthermore, the PIQ provides more school-relevant data, whereas the SDQ examines overall behavior relating to specific problems. The PIQ may therefore be able to provide important insights that are missed by responses to the SDQ alone, and it is especially relevant in school-based studies.

We found that the SDQ and PIQ possessed measurement invariance over time. This confirms a similar finding of measurement invariance over time for the SDQ from DeVries et al.'s (2017). However, we did not find significant changes in any SDQ subscales over grade level, which contrasts with DeVries et al. (2017) finding that these values increased over time. This data reflects self-reports of middle-school students, whereas DeVries et al.'s paper reflected teacher-rated primary school students.

Our inability to find partial measurement invariance across gender contrasts with Bøe et al.'s (2016) finding in a Norwegian sample of adolescents; however, it is in accordance with Van de Looij-Jansen, Goedhart, de Wilde, and Treffer's (2011) results in a Dutch sample. Further work examining measurement invariance for gender in the SDQ is required. Cross-cultural samples would be preferable to eliminate the effect of culture- or language-specific effects.

Several limitations remain in our study. First, we did not explicitly compare different inclusive classroom procedures. A detailed longitudinal comparison may be able to identify the specific effects of different inclusion programs. Second, we tested only the German language version of the PIQ. Lastly, we only used the self-report versions of both tests. More work across multiple language versions, raters, and multiple cultures is still required.

5. Conclusion

Learners with SEN have lower levels of academic self-concept, social inclusion, and emotional inclusion than their typically developing peers; however, these differences shrink dramatically between grades 6 and 7 in inclusive schools. Furthermore, we found that the PIQ scale is valid for making comparisons across gender, grade level, and SEN. The SDQ is similarly valid for grade level and SEN, but not for gender. High values of academic self-concept, social inclusion, and emotional inclusion on the PIQ relate to low levels on the difficulty subscales on the SDQ and high levels of prosocial behavior on the SDQ. We recommend the use of the PIQ for investigations involving inclusive education and for learners with SEN in order to get a detailed picture of how students behave and of how they feel; however, further work relating the PIQ to other instruments and psychological constructs will solidify its usefulness in educational and professional settings.

Acknowledgments

The data presented here was collected as part of a project funded by the Ministry of Education, Science and Culture of Mecklenburg-Vorpommern, Germany, who also approved ethical standards of the study.

References

- Arslan, G. (2018). Exploring the association between school belonging and emotional health among adolescents. *International Journal of Educational Psychology*, 7(1), 21–41.
- Avramidis, E. (2010). Social relationships of pupils with special educational needs in the mainstream primary class: Peer group membership and peer-assessed social behaviour. *European Journal of Special Needs Education*, 25(4), 413–429. <https://doi.org/10.1080/08856257.2010.513550>.
- Avramidis, E., Avgeri, G., & Strogilos, V. (2018). Social participation and friendship quality of students with special educational needs in regular Greek primary schools. *European Journal of Special Needs Education*, 33(2), 221–234. <https://doi.org/10.1080/08856257.2018.1424779>.
- Banks, J., McCoy, S., & Frawley, D. (2017). One of the gang? Peer relations among students with special educational needs in Irish mainstream primary schools. *European Journal of Special Needs Education*, 22(4), 1–16. <https://doi.org/10.1080/08856257.2017.1327397>.
- Bear, G. G., Minke, K. M., & Manning, M. A. (2002). Self-concept of students with learning disabilities: A meta-analysis. *School Psychology Review*, 31(3), 405–427.
- Bøe, T., Hysing, M., Skogen, J. C., & Breivik, K. (2016). The Strengths and Difficulties Questionnaire (SDQ): Factor structure and gender equivalence in Norwegian adolescents. *PLoS One*, 11(5), e0152202. <https://doi.org/10.1371/journal.pone.0152202>.
- Bond, L., Butler, H., Thomas, L., Carlin, J., Glover, S., Bowes, G., ... Patton, G. (2007). Social and school connectedness in early secondary school as predictors of late teenage substance use, mental health, and academic outcomes. *Journal of Adolescent Health*, 40(4), 357. <https://doi.org/10.1016/j.jadohealth.2006.10.013> e9–e18.
- Bossaert, G., Colpin, H., Pijl, S. J., & Petry, K. (2013). Truly included? A literature study focusing on the social dimension of inclusion in education. *International Journal of Inclusive Education*, 17(1), 60–79. <https://doi.org/10.1080/13603116.2011.580464>.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research (Second edition)*. Methodology in the social sciences. New York, London: The Guilford Press.
- DeVries, J. M., Gebhardt, M., & Voß, S. (2017). An assessment of measurement invariance in the 3- and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study. *Personality and Individual Differences*, 119, 1–6. <https://doi.org/10.1016/j.paid.2017.06.026>.
- Dimitrov, D. M. (2017). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121–149. <https://doi.org/10.1177/0748175610373459>.
- Elbaum, B. (2002). The self-concept of students with learning disabilities: A meta-analysis of comparisons across different placements. *Learning Disabilities Research and Practice*, 17(4), 216–226. <https://doi.org/10.1111/1540-5826.00047>.
- Elbaum, B., & Vaughn, S. (2003). For which students with learning disabilities are self-concept interventions effective? *Journal of Learning Disabilities*, 36(2), 101–108.

- <https://doi.org/10.1177/002221940303600203> discussion 149–50.
- Gebhardt, M., Schwab, S., Krammer, M., & Gasteiger, K. (2012). Achievement and integration of students with and without Special Educational Needs (SEN) in the fifth grade. *Journal of Special Education and Rehabilitation*, 13(3–4), <https://doi.org/10.2478/v10215-011-0022-6>.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191. <https://doi.org/10.1007/s10802-010-9434-x>.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, 177(06), 534–539. <https://doi.org/10.1192/bjp.177.6.534>.
- Grütter, J., Gasser, L., & Malti, T. (2017). The role of cross-group friendship and emotions in adolescents' attitudes towards inclusion. *Research in Developmental Disabilities*, 62, 137–147. <https://doi.org/10.1016/j.ridd.2017.01.004>.
- Haeblerlin, U., Moser, U., Bless, G., & Klaghofer, R. (1989). *Integration in die Schulklasse: Fragebogen zur Erfassung von Dimensionen der Integration von Schülern; FDI 4-6; mit einem Heft zur theoretischen und praktischen Einführung einem Testbogen zur Beurteilung des sozialen, emotionalen und leistungsmotivationalen Integriertseins von Schülern des 4. bis 6. Schuljahres in ihre Schulklasse und Auswertungsschablonen [Questionnaire to assess the dimensions of integration of pupils; FDI 4-6; an assessment of the social, emotional, and motivational integration of 4th-6th grade students, including a theoretical and practical introduction and evaluation templates]. Beiträge zur Heil- und Sonderpädagogik, Vol. 8.* Bern u.a: Haupt.
- Hagquist, C. (2007). The psychometric properties of the self-reported SDQ – An analysis of Swedish data based on the Rasch model. *Personality and Individual Differences*, 43(5), 1289–1301. <https://doi.org/10.1016/j.paid.2007.03.022>.
- Hu, L.-L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Huber, C., Gerullis, A., Gebhardt, M., & Schwab, S. (2018). The impact of social referencing on social acceptance of children with disabilities and migrant background: An experimental study in primary school settings. *European Journal of Special Needs Education*, 33(2), 269–285. <https://doi.org/10.1080/08856257.2018.1424778>.
- Korhonen, J., Linnanmäki, K., & Aunio, P. (2014). Learning difficulties, academic well-being and educational dropout: A person-centred approach. *Learning and Individual Differences*, 31, 1–10. <https://doi.org/10.1016/j.lindif.2013.12.011>.
- Levine, D. W., Kaplan, R. M., Kripke, D. F., Bowen, D. J., Naughton, M. J., & Shumaker, S. A. (2003). Factor structure and measurement invariance of the women's health initiative insomnia rating scale. *Psychological Assessment*, 15(2), 123–136. <https://doi.org/10.1037/1040-3590.15.2.123>.
- Li, E. P.-Y., Tam, A. S.-F., & Man, D. W.-K. (2006). Exploring the self-concepts of persons with intellectual disabilities. *Journal of Intellectual Disabilities: JOID*, 10(1), 19–34. <https://doi.org/10.1177/17446295060062270>.
- McGraw, K., Moore, S., Fuller, A., & Bates, G. (2008). Family, peer and school connectedness in final year secondary school students. *Australian Psychologist*, 43(1), 27–37. <https://doi.org/10.1080/00050060701668637>.
- Mellins, C. A., Xu, Q., Nestadt, D. F., Knox, J., Kauchali, S., Arpadi, S., ... Davidson, L. L. (2018). Screening for mental health among young South African children: The use of the Strengths and Difficulties Questionnaire (SDQ). *Global Social Welfare*, 5(1), 29–38. <https://doi.org/10.1007/s40609-018-0110-3>.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's guide* (seventh edition). (1998–2015) Los Angeles, CA: Muthén & Muthén.
- Ortuño-Sierra, J., Chocarro, E., Fonseca-Pedrero, E., Riba, S. S., & Muñiz, J. (2015). The assessment of emotional and behavioural problems: Internal structure of the Strengths and Difficulties Questionnaire. *International Journal of Clinical and Health Psychology*, 15(3), 265–273. <https://doi.org/10.1016/j.ijchp.2015.05.005>.
- Ortuño-Sierra, J., Fonseca-Pedrero, E., Aritio-Solana, R., Velasco, A. M., Luis, E. C., de Schumann, G., ... Lawrence, C. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European Child & Adolescent Psychiatry*, 24(12), 1523–1534. <https://doi.org/10.1007/s00787-015-0729-x>.
- Prince, E. J., & Hadwin, J. (2013). The role of a sense of school belonging in understanding the effectiveness of inclusion of children with special educational needs. *International Journal of Inclusive Education*, 17(3), 238–262. <https://doi.org/10.1080/13603116.2012.676081>.
- Rose, C. A., Monda-Amaya, L. E., & Espelage, D. L. (2010). Bullying perpetration and victimization in special education: A review of the literature. *Remedial and Special Education*, 32(2), 114–130. <https://doi.org/10.1177/0741932510361247>.
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75(2), 243–248. <https://doi.org/10.1007/s11336-009-9135-y>.
- Sauer, S., Ide, S., & Borchert, J. (2007). Zum Selbstkonzept von Schülerinnen und Schülern an Förderschulen und in integrativer Beschulung: Eine Vergleichsuntersuchung [The self-concept of students in special and integrated schools: a comparative study]. *Heilpädagogische Forschung*, 33(3), 135–142.
- Schwab, S. (2017). The impact of contact on students' attitudes towards peers with disabilities. *Research in Developmental Disabilities*, 62, 160–165. <https://doi.org/10.1016/j.ridd.2017.01.015>.
- Schwab, S., Gebhardt, M., & Gasteiger-Klicpera, B. (2013). Predictors of social inclusion of students with and without SEN in integrated settings. *Croatian Review of Rehabilitation Research*, 49(Supplement), 106–114.
- Schwab, S., Gebhardt, M., Krammer, M., & Gasteiger-Klicpera, B. (2015). Linking self-rated social inclusion to social behaviour. An empirical study of students with and without special education needs in secondary schools. *European Journal of Special Needs Education*, 30(1), 1–14. <https://doi.org/10.1080/08856257.2014.933550>.
- Stiefel, L., Shiferaw, M., Schwartz, A. E., & Gottfried, M. (2017). Who feels included in school? Examining feelings of inclusion among students with disabilities. *Educational Researcher*. <https://doi.org/10.3102/0013189X17738761> 0013189X1773876.
- 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. <https://doi.org/10.1016/j.lindif.2015.08.037>.
- United Nations Convention on the Rights of Persons with Disabilities, 2006. United Nations.
- Van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. A. (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *The British Journal of Clinical Psychology*, 50(2), 127–144. <https://doi.org/10.1348/014466510X498174>.
- Van Ryzin, M. J., Gravelly, A. A., & Roseth, C. J. (2009). Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being. *Journal of Youth and Adolescence*, 38(1), 1–12. <https://doi.org/10.1007/s10964-007-9257-4>.
- Venet, M., Zurbriggen, C., & Eckhart, M. (2014). Entwicklung und erste Validierung einer Kurzversion des "Fragebogens zur Erfassung von Dimensionen der Integration von Schülern (FDI 4-6)" von Haeblerlin, Moser, Bless und Klaghofer [Development and first validation of a short form of the Questionnaire to assess the dimensions of integration of pupils (FDI 4-6) from Haeblerlin, Moser, Bless und Klaghofer]. *Empirische Sonderpädagogik*, 2, 99–113.
- Venet, M., Zurbriggen, C., Eckhart, M., Schwab, S., & Hessels, M. G. P. (2015). *The perceptions of inclusion questionnaire (PIQ). English version*. Retrieved from www.piqinfo.ch.
- Voß, S., Hauer, A., Blumenthal, Y., Mahlau, K., Sikora, S., & Hartke, B. (2017). *Zum Leistungs- und Entwicklungsstand inklusiv beschulter Schülerinnen und Schüler mit (sonder-)pädagogischen Förderbedarfen auf der Insel Rügen nach sechs Schulbesuchsjahren [Evaluation and state of development for inclusion of students with special education needs on the island of Rügen after six years of schooling]*. Retrieved from https://www.rim.uni-rostock.de/fileadmin/uni-rostock/Alle_PHF/RIM/Downloads/RIM-Evaluationsbericht-MZP7_Internet.pdf.
- Wei, X., & Marder, C. (2010). Self-concept development of students with disabilities. *Remedial and Special Education*, 33(4), 247–257. <https://doi.org/10.1177/0741932510394872>.
- Wight, M., & Chapparo, C. (2008). Social competence and learning difficulties: Teacher perceptions. *Australian Occupational Therapy Journal*, 55(4), 256–265. <https://doi.org/10.1111/j.1440-1630.2007.00706.x>.
- Zurbriggen, C. L. A., Venet, M., Schwab, S., & Hessels, M. G. P. (2017). A psychometric analysis of the student version of the perceptions of inclusion questionnaire (PIQ). *European Journal of Psychological Assessment*, 1–9. <https://doi.org/10.1027/1015-5759/a000443>.