Academic resilience: underlying norms and validity of definitions

Wangqiong Ye, et al. *[full author details at the end of the article]*

Received: 16 January 2020 / Accepted: 10 December 2020 / Published online: 19 January 2021
© The Author(s) 2021

Abstract

Academic resilience refers to students' capacity to perform highly despite a disadvantaged background. Although most studies using international large-scale assessment (ILSA) data defined academic resilience with two criteria, student background and achievement, their conceptualizations and operationalizations varied substantially. In a systematic review, we identified 20 ILSA studies applying different criteria, different approaches to setting thresholds (the same fixed ones across countries or relative country-specific ones), and different threshold levels. Our study on the validity of these differences and how they affected the composition of academically resilient students revealed that the classification depended heavily on the threshold applied. When a fixed background threshold was applied, the classification was likely to be affected by the developmental state of a country. This could result in an overestimation of the proportions of academically resilient students in some countries while an underestimation in others. Furthermore, compared to the application of a social or economic capital indication, applying a cultural capital indicator may lead to lower shares of disadvantaged students classified as academically resilient. The composition of academically resilient students varied significantly by gender and language depending on which indicator of human capital or which thresholds were applied reflecting underlying societal characteristics. Conclusions drawn from such different results depending on the specific conceptualizations and operationalizations would vary greatly. Finally, our study utilizing PISA 2015 data from three countries representing diverse cultures and performance levels revealed that a stronger sense of belonging to a school significantly increased the chances to be classified as academically resilient in Peru, but not in Norway or Hong Kong. In contrast, absence from school was significantly associated with academic resilience in Norway and Hong Kong, but not in Peru.

Keywords International large-scale assessments · Educational inequality · Human capital · Student achievement · Socio-economic background · Systematic review

1 Introduction

Resilience refers to successful adaption to situations despite risks that put someone at a disadvantage or adversity (Ungar 2005; Windle et al. 2011). In line with this general definition, academic resilience refers to the capacity of students to perform well in

school despite a disadvantaged background (OECD 2011) or more precisely the heightened likelihood of success in school despite environmental adversities brought about by early traits, conditions, and experiences (Wang et al. 1994).

Since minimizing the influence of students' background on the outcomes of schooling is a central topic for accomplishing equity in education, a better understanding of academic resilience may help policymakers and educators to support students from a disadvantaged background in improving their academic performance. However, different conceptualizations of academic resilience may result in conflicting conclusions. It is therefore crucial to ensure the validity of a definition.

Studies on academic resilience typically employ some operationalization of socio-economic status (SES) as an indicator of students' risk or adversity, and they use some type of educational outcome as an indicator of positive adaptation (Tudor and Spray 2017). Thresholds are usually used to combine continuous SES and outcome measures into a binary variable that indicates academic resilience or non-resilience.

In the context of international large-scale assessments (ILSAs), most studies adopted a composite SES index to operationalize students' background. General problems such as missing data or questionable comparability of this index across countries (Watermann et al. 2016), are specifically related to the conceptualization of academic resilience: A composite SES index treats student background as one-dimensional. Thus analyses based on such an index do not reveal the potential relevance of different SES components. Furthermore, studies applied different thresholds to define a disadvantaged background (and also to what it means to perform well). Whereas some studies used the same fixed thresholds for all countries included in their study, others used relative thresholds derived from the data within each country. The evidence supporting the validity of these decisions was often quite limited.

Since the measurement of academic resilience is inherently influenced by definitional issues, this study sought to examine the validity of different conceptualizations of academic resilience and how these affect the composition of academically resilient students. For this purpose, three countries were selected representing diverse cultures and performance levels (Norway, Peru, and Hong Kong). Student performance in science was used as an indicator of educational outcomes.

Besides the common composite SES index also used in other studies, three specific background indicators representing different dimensions of SES (economic, cultural, and social) were adopted to operationalize student background. Two types of thresholds (the same fixed and relative within-country thresholds) were applied to define a disadvantaged student background or high performance. Thus, in total, sixteen conceptualizations of academic resilience were examined on their validity, with four background indicators and two types of thresholds.

To illustrate how many and which students were classified as academically resilient, we selected two individual student characteristics (gender and language spoken at home). As validity measures, we selected two school-related characteristics (sense of belonging and absence from school) that can be supposed to assess similar concepts. This study examined their concurrent validity by comparing the relations of these external constructs to the different conceptualizations of academic resilience.

The paper is organized as follows. Firstly, a conceptual framework is developed that distinguishes between different ways to define academic resilience, including their underlying norms. Secondly, an overview of the literature about academic resilience is provided,

in particular in the context of ILSAs. Research gaps and the research questions examined in this paper are presented thereafter. Thirdly, a methods section follows that provides information about the data and variables used and the analyses applied, results are presented after that. Finally, the paper concludes with a summary and a discussion of implications.

1.1 Conceptual framework: criteria of academic resilience and underlying norms

1.1.1 Resilience and academic resilience

Research on resilience in the behavioral sciences began to emerge around 1970. Since the mid-1980s, an increasing number of researchers from different disciplines (e.g., child development, pediatrics, psychology, psychiatry, and sociology) have published findings from studies on children who were successful in life despite adverse childhood environments (Werner 2000). The theoretical development about resilience has went through four waves: (1) identifying resilient qualities, (2) uncovering the resilience process, (3) promoting resilience through prevention and intervention, and (4) focusing on the dynamics of adaptation and change (Masten 2007). The latter means that resilience may vary across contexts and over time (Tudor and Spray 2017).

Although there is no universal definition for resilience across the different disciplines examining this phenomenon, most definitions are based around two core concepts: adversity and positive adaptation (Windle 2011). Correspondingly, in the context of schooling, academic resilience is defined by some measure of adversity in terms of early traits, conditions and experiences and by some measure of increased likelihood to succeed in school (Wang et al. 1994).

1.1.2 Measuring adversity: composite vs. distinct measures of student background

From a theoretical point of view, it is possible to distinguish between different dimensions (e.g., education, social status, and wealth) of an individual's background that may predefine his or her chances later in life. In major theories, the effects of social background on student outcomes are therefore conceptualized not only as a consequence of material possessions but also based on social and cultural practices (Bourdieu 1986). According to Bourdieu's capital theory (1986), individuals process economic, cultural, and social capital such as monetary resources, cultural possessions, and social relationships. These three types of capital can be distinguished, and each of them can be used for the accumulation of other types of capital.

Academic resilience studies typically use a composite index that covers several of these background dimensions. For example, the composite SES index of the PISA studies, economic, social, and cultural status (ESCS), includes parents' occupation, parents' education, and home resources (OECD 2017). Although the ESCS covers two of the three Bourdieu dimensions of capital, it is treated as a one-dimensional measure. Consequently, analyses conducted with this index cannot reveal the relevance of the different SES subdimensions for being academically resilient.

Studies on academic resilience using International Association for the Evaluation of Educational Achievement (IEA) data, for example TIMSS 2015, usually use the Home Educational Resources (HER) index, which is based on parents' education, the number of books at home, and home study support (Mullis and Martin 2013). It is therefore

mostly a measure of students' cultural capital. Parents' occupation status as an indicator of students' economic capital was not included in the HER index, but was a part of another SES index for Grade four students, Home Resources for Learning (HRL).

Since the measurement of academic resilience is inherently influenced by conceptual issues (Windle et al. 2011), including alternative measures of social, cultural, and economic capital in the definition may shed light on how these dimensions of social background affect the results (Watermann et al. 2016). The present study follows this idea and assesses adversity with both composite and distinct measures.

A specific challenge is that the differences between countries make it challenging to use the same background measures to study academic resilience across countries (Coronado-Hijón 2017). For example, owning a car is often used as one indicator of student background, but this may have different meanings in economically developed and developing countries. Some of the measures used in the present study address this challenge; we will examine this issue further in the discussion.

1.1.3 Measuring positive adaptation: selecting an Indicator of student outcome

Educational outcomes can be distinguished into cognitive and non-cognitive (Heckman et al. 2006). Unlike resilience studies in psychology, non-cognitive outcomes were rarely used to measure positive adaptation in education (Tudor and Spray 2017). Non-cognitive skills like self-efficacy or educational aspiration were merely regarded as protective factors promoting academic resilience, or as outcomes of being resilient (OECD 2018). As a result, these studies tended to use cognitive outcomes, especially test scores to measure positive adaptation.

Test scores stem either from one or several subject domains. In case of using one subject domain, most studies focused on reading, mathematics, or science. These domains were regarded as providing fundamental skills needed for further education or success in the labor market (OECD 2018). Thus, one purpose of these studies was to shed light on the competitiveness of a country.

Considering that positive adaptation may vary by domain, some studies used data from different domains. OECD (2011) found students who showed positive adaptation in science did usually so also in mathematics or/and reading. However, other studies found that positive adaptation in one domain was not necessarily associated with positive adaptation in other domains. Therefore, they defined resilience as a characteristic across domains, for example, by showing positive adaptation in reading, mathematics, and sciences (Agasisti et al. 2018).

As previously stated, studies using cognitive skills to operationalize positive adaptation usually treated traits like anxiety, motivation, or engagement as predictors or outcomes associated with resilience due to bidirectional developmental processes (Coronado-Hijón 2017). Therefore, some educational researchers recently also began to use non-cognitive outcomes to assess positive adaptation in resilience studies (OECD 2018).

1.1.4 Thresholds for adversity and positive adaptation: cross-country vs. within-country

Despite decisions on selecting indicators of adversity and positive adaptation, another step in conceptualizing academic resilience is to decide about the thresholds, which

define a “disadvantaged” background (adversity) or “high” performance (positive adaptation). These decisions vary substantially across studies. One core distinction is between “fixed” and “relative” thresholds. “Fixed” means that the same threshold is applied across countries, whereas “relative” means that based on within-country data, different thresholds are used for different countries.

Using fixed thresholds stresses an international perspective where direct cross-country comparisons are at the forefront. In this perspective, the proportion of the resilient student is regarded as an indicator for quality and equity of education systems (Erberer et al. 2015; OECD 2011). Using relative thresholds means to define academic resilience from a national perspective, provides important insights on policy levers that are associated with resilience within different education systems (OECD 2011). When relative thresholds were applied, for example, successful disadvantaged students in one country may be classified as poor performing in other contexts.

A similar distinction as the one between fixed and relative thresholds is frequently made in the research on poverty that differentiates between absolute and relative poverty (Hagenaars and De Vos 1988). Research on academic resilience is more complex because it combines information from two criteria, student background and educational outcome. Therefore, we need to distinguish between four possible approaches to define academic resilience: (1) a fixed threshold for background and a fixed threshold for outcome; (2) a fixed threshold for background but a relative threshold for outcome; (3) a relative threshold for background but a fixed threshold for outcome; and (4) a relative threshold for background and a relative threshold for outcome. Several cutoff values (e.g. 20%, 25%, or 33%) were used to define thresholds in many studies; considering the economic and performance differences among our three samples, cutoff value 33% was adopted to have more students for analysis. Details are reported below.

1.2 State of research

1.2.1 Overview about academic resilience studies in international large-scale assessments

Since ILSAs have facilitated cross-country analyses of student achievement and its predictors, there is an increasing number of studies using data from ILSAs to investigate how individual and institutional features are related to academic resilience (Gonzalez and Padilla 1997; Martin and Marsh 2006; Sandoval-Hernández and Bialowolski 2016). To summarize the state of research, a systematic literature search in Web of Science, ERIC, and Google Scholar was carried out in July 2019. The search was built around four groups of key words: education (e.g., academic), resilience (e.g., resilient, buoyance), measurement (e.g., scale), and ILSA (e.g., PISA). The search was limited to English-language publications and revealed about 20 studies directly related to our topic (see Table 1). They applied a broad range of different criteria, different approaches to setting thresholds, and different threshold levels.

Table 1 shows the different operationalizations, which were grouped according to the four approaches to set thresholds explained above. As the overview reveals, most studies used Organization for Economic Co-operation and Development (OECD) data rather than IEA data. One reason could be the tremendous influence of PISA (Meyer

Table 1 Operationalizations of academic resilience in ILSAs

Type of threshold	Reference	Data	Student background indicator (threshold for defining adversity)	Outcome (threshold for defining positive adaptation)	Domain, Country samples	Main findings
Fixed background, fixed outcome	Erberer et al. (2015)	TIMSS 2011	HER index (“few resources”)	Fix (≥ 475)	Mathematics (Grade 8), 28 education systems	Environments of high academic achievement appear to support academic resilience. Students’ high educational aspirations appear to be the strongest and most consistent predictor of academic resilience.
	Sandoval-Hernández and Bialowski (2016)	TIMSS 2011	HER index (“few resources”)	Fix (≥ 475)	Mathematics (Grade 8), 5 Asian education systems	No consistent patterns were identified, student expectations and amount of time spent on homework were significantly associated with academic resilience in Singapore, boys tended to be more resilient in Korea than girls.
	Frempong et al. (2016)	TIMSS 2012	New SES index based on 18 assets listed in student questionnaire: disadvantaged students (bottom 2 SES quantiles) from disadvantaged schools (bottom 3 quantiles)	Fix (≥ 352 as the South African country mean)	Mathematics (Grade 9), South Africa	Girls who speak their native language tended to be more resilient. Positive attitude, valuing learning, job aspiration were important characteristics associated with academic resilience.
Fixed background, relative outcome	Sandoval-Hernandez and Cortés (2012)	PIRLS 2006	new SES index based on parents’ education, occupation, home possessions (bottom 20% across countries in a cluster)	Within country (top 20%)	Reading, cluster of countries with comparable SES index	Non-cognitive characteristics (self-confidence and motivation) were the most important factors in predicting academic resilience. Material support provided by family and safety in school also mattered.
Relative background, fixed outcome	OECD (2011)	PISA 2009	ESCS (bottom 1/3 within country)	Across countries (top 1/3)	Science all participating countries	Academic resilience was not identified as a domain-specific characteristic. No gender gap in academic resilience was found for science. Language and

Table 1 (continued)

Type of threshold	Reference	Data	Student background indicator (threshold for defining adversity)	Outcome (threshold for defining positive adaptation)	Domain, Country samples	Main findings
	OECD (2017)	PISA 2015 PISA 2006	ESCS (bottom 1/4 within country)	Across countries (top 1/4)	Science, all participating countries	immigrant background were marginally associated with academic resilience in a few countries. In countries with a large increase in the proportion of academically resilient students, the percentage of low-performing students was reduced while the average performance was maintained or improved.
	OECD (2018)	PISA 2015	ESCS (bottom 1/4 within country)	Across countries (top 1/4)	Science, all participating countries	The presence of academic resilience varies greatly in this operationalization, 76% of disadvantaged students are academically resilient in Viet Nam, but less than 5% in Kosovo, Peru, and Tunisia.
	García-Crespo et al. (2019)	PIRLS 2016	New SES index based on possessions and books at home, highest academic qualifications, and level of employment of parents (bottom 25% within countries)	Across EU countries (top 25%)	Reading, EU countries that participated in this study	Student confidence in reading and a favorable school climate greatly increased the likelihood of academic resilience.
Residual	OECD (2010)	PISA 2009	ESCS (bottom 1/4 within country)	Across countries (top 1/4 of residual)	Reading, all participating countries	Academically resilient students were more prevalent in those education systems that PISA indicators showed to be more equitable.
	Cheung et al. (2014)	PISA 2009	ESCS (bottom 1/4 within country)	Across countries (top 1/4 of residual)	Reading, 4 East Asian economies	Family structure, expected education, kindergarten attendance, and reading engagement were associated with academic resilience.

Table 1 (continued)

Type of threshold	Reference	Data	Student background indicator (threshold for defining adversity)	Outcome (threshold for defining positive adaptation)	Domain, Country samples	Main findings
	Agasisti and Longobardi (2014)	PISA 2009	School ESCS and student SES (bottom 1/3 students from bottom 1/3 schools within country)	Across countries (top 1/3 of residual)	Reading, Italy	Motivation and extracurricular activities were positively associated with academic resilience. Immigrant status was negatively associated with academic resilience.
	Agasisti and Longobardi (2014)	PISA 2000–2012	School ESCS and student SES (bottom 1/3 students from bottom 1/3 schools within country)	Across countries (top 1/3 of residual)	Reading, 58 education systems over period 2000–2012	Educational funding can help disadvantaged students in overcoming their penalizing starting conditions.
	Cheung (2017)	PISA 2012	ESCS (bottom 1/4 within country)	Across countries (top 1/4 of residual)	Mathematics, Five East Asian education systems	Self-efficacy was strongly, familiarity with mathematical concepts learned in the earlier grades moderately, and anxiety weakly associated with academic resilience.
	Agasisti and Longobardi (2017)	PISA 2009	School ESCS and student SES (bottom 1/3 students from bottom 1/3 schools within country)	Across countries (top 1/3 of residual)	Reading, 5 EU countries	More extracurricular activities and a more positive school climate were positively associated with academic resilience.
Across-domain	Agasisti et al. (2018)	PISA 2006–2015	ESCS (bottom 1/4 within country)	≥ “Level 3” in all three domains	Reading, mathematics and science, all participating countries	Classroom disciplinary climate was positively associated with academic resilience. The amount of human and material resources available in school was weakly associated with academic resilience.
	OECD (2018)	PISA 2015	ESCS (bottom 1/4 within country)	≥ “Level 3” in all three domains	Reading, mathematics and science, all participating countries	This type of cross-domain academic resilience was more frequently observed in countries with higher average performance. Disciplinary climate and motivation were associated with academic resilience.

Table 1 (continued)

Type of threshold	Reference	Data	Student background indicator (threshold for defining adversity)	Outcome (threshold for defining positive adaptation)	Domain, Country samples	Main findings
Non-cognitive	OECD (2018)	PISA 2015	ESCS (bottom 1/4 within country)	Socially and emotionally satisfied	Non-cognitive, 48 countries	Students who were socially and emotionally resilient also tended to do better academically. Lower shares of socio-emotionally resilient students were found in the top-performing systems than with the application of a cognitive outcome definition of resilience.
Relative background, relative outcome	OECD (2011)	PISA 2009	ESCS (bottom 1/3 within country)	Within country (top1/3)	Science, all participating countries	The extent to which students adopted positive approaches to learning and the amount of time they spent in regular science lessons were strongly associated with academic resilience.
	Karklina (2012)	PISA 2006 interview data	ESCS (bottom 1/3 within country)	Within country (top1/3)	Science, Latvia	Parental education level and length of poverty experience were associated with academic resilience.
	Aydiner and Kalender (2015)	PISA 2012	ESCS, bottom 1/4 within country	Within country (top1/3)	Reading, Turkey	Sense of belonging was associated with academic resilience
	OECD (2018)	PISA 2015	ESCS, bottom 1/4 within country	top1/4 within countries	Science, all participating countries	Academic resilience was not more frequently observed in top-performing countries than in developing or low-performing countries.

HER Home Educational Resources Scale; *ESCS* economic, social, and cultural status index; *School ESCS* average student ESCS in a school; *SES* socio-economic status; *Level 3* score higher than 484.14 and less than or equal to 558.73; *socially and emotionally satisfied* students satisfied with their life, felt socially integrated at school and did not suffer from test anxiety

et al. 2017), another possible reason could be the missing data problem on student background indicators in IEA data (Broer et al. 2019). Therefore, studies using IEA data to explore academic resilience often either adopted a self-developed SES index (García-Crespo et al. 2019) or focused on selected countries with enough SES information (Cheung 2017; Erberer et al. 2015).

We will next review the ILSA research on academic resilience with respect to the conceptualizations and operationalizations used (see Table 1). For substantive results of these studies, please see the last column in this table.

1.2.2 Fixed background and fixed outcome thresholds

We identified three studies that used fixed thresholds to define both disadvantage and positive adaptation across different countries. Erberer et al. (2015) examined how prevalent academic resilience was across education systems and which protective factors could be identified. Their study used TIMSS 2011 data and adopted the composite Home Educational Resources (HER) index as a family SES measure. The authors classified a student as disadvantaged by applying a fixed threshold (a score ≤ 7.3 on the HER scale). Meanwhile, the authors used the so-called TIMSS International Intermediate Benchmark of Mathematics (students that reached this benchmark can apply basic mathematical knowledge in simple situations) as a threshold (a score ≥ 475) to define positive adaptation.

Sandoval-Hernández and Bialowolski (2016) adopted Erberer et al.'s (2015) method, and applied the definition to TIMSS 2011 data from five Asian education systems. Frempong et al. (2016) also followed this procedure and applied the definition to TIMSS 2011 data from South Africa. Frempong et al. did not adopt the HER index but calculated student SES index based on 18 assets listed in the student questionnaire.

In these three studies, the fixed thresholds for achievement to define positive adaptation were set either around the international (Erberer et al. 2015; Sandoval-Hernández and Bialowolski 2016) or the national mean (Frempong et al. 2016).

1.2.3 Fixed background and relative outcome thresholds

Our systematic review revealed only one study that adopted a fixed threshold to define adversity and a relative threshold to define positive adaptation. Sandoval-Hernández and Cortés (2012) applied the concept of academic resilience to Progress in International Reading Literacy Study (PIRLS) 2006 data. Since PIRLS 2006 does not provide a composite SES index (Mullis et al. 2004), authors followed Caro and Cortés's method (2012) and calculated an index based on parents' education, parents' occupation status, and home possessions. Considering measurement invariance, authors restricted their analysis to a cluster of countries with a comparable SES index. Disadvantaged background was defined by adopting a fixed SES threshold which was the 20th percentile of the index in the pooled data of all countries in the cluster. Positive adaptation was defined by a relative threshold which was the 80th percentile in each country.

1.2.4 Relative background and fixed outcome thresholds

Within this approach to define academic resilience, a methodological difference was found how to use the thresholds set. These were either used directly as in the studies

described above or each disadvantaged student's performance was compared with the performance predicted by the average relationship among students from similar SES backgrounds across countries. The difference between these two was called a student's "residual" performance. Furthermore, within this group of studies, one of them used non-cognitive skills as an indicator of educational outcomes, and two of them used an across-domain operationalization of educational outcomes.

Direct threshold approaches OECD (2011) adopted the composite index ESCS and defined disadvantaged students by a relative background threshold (bottom 1/3 of ESCS within each country), whereas positive adaptation was defined by a fixed threshold (top 1/3 of students' performance across countries). OECD (2017) narrowed both thresholds down by defining academically resilient students as those who were in the bottom 1/4 of ESCS within each country and performed in the top 1/4 of students across all participating education systems. OECD (2018) adopted the same operationalization.

García-Crespo et al. (2019) explored predicting factors of academic resilience in reading literacy at Grade four, using PIRLS 2016 data from European Union member countries. The authors calculated their own Social, Economic, and Cultural Index (SECI) to measure student SES, based on home possession, number of books in the home, the highest academic qualifications of the parents, and the highest level of employment of the parents. Students in the bottom 25% of the SECI within each country, with a performance in the top 25% across the participating EU countries, were considered to be academically resilient.

Residual methods to calculate thresholds OECD (2010) defined disadvantaged students as those in the bottom 1/4 of ESCS within each country, while disadvantaged students in the top 1/4 of residual performance across countries were classified as academically resilient.

Several studies adopted this residual method, although OECD (2011) itself adopted new methods in its later studies (OECD 2018). Cheung et al. (2014) applied the residual method to PISA 2009 data from four East Asian economies in reading literacy. Academically resilient students were defined as those in the bottom 1/4 of ESCS within each country who achieved the top 1/4 residual performance across countries. Cheung (2017) applied the same definition to PISA 2012 data and examined academic resilience in mathematics, and also focused on a cluster of East Asian education systems.

Agasisti and Longobardi (2014, 2017) put special emphasis on disadvantaged students in disadvantaged schools and applied their definition to a group of European countries. The authors firstly selected schools among the 1/3 bottom of ESCS within each country based on the aggregated school ESCS average. From these schools, they selected those students who were in the 1/3 bottom of ESCS within the country. Resilient students were defined as disadvantaged students from disadvantaged schools who have a residual performance among the top 1/3 across countries.

Studies in this group usually focused on a cluster of countries with comparable economic and cultural backgrounds, because the strength between SES and performance varied across countries.

Across-domain operationalization of educational outcomes The studies mentioned above included only one domain (reading, science, or mathematics) as an indicator of positive adaptation. Agasisti et al. (2018) were the first to examine academic resilience across the three core domains in PISA—reading, mathematics, and science. Academically resilient students were defined as those among the bottom 1/4 of ESCS within each country, who performed at or above Proficiency Level 3 (i.e., one above the baseline level of proficiency needed to participate in society) in all three PISA domains. OECD (2018) adopted the same operationalization.

Outcome definition including non-cognitive characteristics Most studies in the ILSA context used cognitive outcomes (e.g., school achievement) to define positive adaptation, whereas non-cognitive skills (e.g., motivation) were treated as protective factors rather than indicators of positive adaptation. OECD (2018) examined for the first time non-cognitive outcomes and defined resilience in a non-cognitive way. Disadvantaged students from the bottom 1/4 of the ESCS distribution within each country were considered to be “socially and emotionally resilient”, if they were satisfied with their life, felt socially integrated at school, and did not suffer from test anxiety (OECD 2018). When this definition was applied, lower shares of resilient students were found in the top-performing Asian educational systems than with the application of a cognitive outcome definition.

1.2.5 Relative background and relative outcome thresholds

Our systematic review identified four studies applying relative thresholds to defining both adversity and positive adaptations. As OECD (2011) mentioned, the purpose was to support policy makers and stakeholders with knowledge about how to foster resilience within their education systems. Disadvantaged students were defined by a relative threshold (bottom 1/3 of ESCS within each country), and the threshold for performance was also set as a relative one (top 1/3 within each country). Karklina (2012) used the same operationalization with PISA 2006 data from Latvia. Aydiner and Kalender (2015) adopted this approach as well but changed the cutoff values for thresholds—bottom 1/4 ESCS within a country for the background threshold and top 1/4 among the disadvantaged students within a country for the performance threshold. OECD (2018) followed this approach in its study about resilience from a national perspective and classified students from the bottom 1/4 of the ESCS distribution within each country and a performance among the top 1/4 of science within each country as resilient.

1.3 Does the conceptualization of academic resilience matter: a question of validity

In summary, validity may be defined as the extent to which we can back up the inferences drawn from an assessment by arguments based on evidence (Kane et al. 2005). The present study investigates the criterion validity of different conceptualizations of academic resilience, which means their relation to external criteria. Concurrent validity, where both the actual construct and the criterion measures are supposed to assess the same underlying trait and are collected at the same time, is a core dimension of criterion validity (Cohen and Swerdlik 2018).

Concurrent validity is demonstrated when a measure is positively or negatively correlated with another relevant measure as hypothesized, or when a new measure is associated with one that was already considered valid (Fink 2010). Two external criteria that should be strongly associated with academic resilience were applied, namely sense of belonging (positively) and absence from class (negatively), both of which have been identified in the literature as predictors of academic resilience (Sandoval-Hernández and Bialowolski 2016; Tommaso et al. 2018). A sense of belonging influences student outcomes via its effects on motivation and engagement, which were considered predictors of academic resilience in many studies (Aydiner and Kalender 2015; OECD 2011). Similarly, studies revealed that students who did not frequently skip class were more likely to be resilient (OECD 2018). The purpose of our study is to examine whether the strength of these relations varied by conceptualization of resilience.

Furthermore, we applied two background characteristics often used in the literature to describe the groups of students classified as academically resilient, namely gender and the language spoken at home (Cheung et al. 2014; OECD 2011) with the purpose to see how the group compositions changes depending on the conceptualization.

The aim of our study is to examine how different conceptualizations of academic resilience affect which students (gender and language) are classified as resilient, and to what extent the conceptualizations correspond with the two external criteria (sense of belonging and absence from school). Furthermore, given that the operationalization of student background may be affected by cultural differences and student performance varies substantially across countries, we examined concurrent validity for countries representing different cultures and performance levels.

2 The present study

As illustrated, four different approaches were applied to conceptualize academic resilience, either by using fixed or relative thresholds with respect to defining a disadvantaged student background or a strong educational outcome (positive adaptation). How the different approaches work empirically is largely an open question. For example, applying the same fixed thresholds to indicators of students' adversity may not work well in both developing and developed countries, because they may provide either very large or small groups of students classified as disadvantaged just because the whole country is less or more developed than others. Similarly, applying the same fixed threshold to indicators of positive adaptation may not work well in both high-performing and low-performing countries because they may lead to either very large or small groups of students classified as high-achieving just because the whole country performs more or less well. Furthermore, where thresholds were set varied substantially across studies. Little is known what these differences may mean regarding their relation to external criteria.

In addition, studies using PISA data to examine academic resilience can make use of a composite measure for SES (OECD 2005). However, the availability of this ESCS index is a double-edged sword to resilient studies because adopting the index without providing validity evidence or examining its subdimensions may underestimate their relevance (Watermann et al. 2016).

This study aims to examine the validity of different conceptualizations and operationalizations of academic resilience with data from countries that represent different developmental and achievement levels. Hong Kong was a high-achieving country (9th out of 72 participating countries), Norway was above average (24th), and Peru was near the bottom (66th) in the 2015 PISA cycle. Regarding the developmental status of these countries, Norway and Hong Kong both ranked highly on the human development index while Peru ranked low. It is worthwhile to mention that the rank of Hong Kong dropped dramatically in the inequality adjusted Human Development Index (UNDP 2016). Compared to other economies, Hong Kong has a relatively high-income disparity (Hong Kong Economy 2010). However, the relationship between SES and mathematics achievement was found to be the lowest among participating economies in PISA 2012 (Kalaycıoğlu 2015), which suggested high educational quality and equity in the system. Norway has a reputation for equity in its education system (Reimer et al. 2018), and empirical studies have found a pronounced increase of academic resilience in Norway from 2006 to 2015 (Agasisti et al. 2018). At the opposite extreme, empirical studies also found an extremely low percentage of resilient students in Peru. The three countries are in addition geographically separated and represent very different cultures.

Against this background, our study aims at answering the following questions:

1. How large is the group of academically resilient students when different conceptualizations of academic resilience are applied?
2. How do these conceptualizations of academic resilience affect which students are classified as academically resilient when it comes to gender and language background?
3. How are different conceptualizations of academic resilience associated with external variables, which can be supposed to assess similar constructs?
4. Do results change if different indicators of students' capital (economic, social, and cultural) are used?

3 Methods

3.1 Sample

This study used data from PISA 2015, which covered science, reading, mathematics, and financial literacy. Science was selected because it was the primary focus of this cycle and thus provided more precise estimates than for the other domains. Given the assumed relevance of a country's developmental and achievement state, this study used information from three education systems representing different economic contexts and performance levels—Hong Kong, Norway, and Peru.

PISA uses a two-stage stratified sampling strategy. Schools are sampled in the first stage, with the probability of selection being dependent on the number of eligible students enrolled. In the second stage, a random sample of students, aged from 15 years and 3 months to 16 years and 2 months, is selected within schools. Our total sample included 239 schools with an average about 24 students in Norway, 282 schools with an average 25 students in Peru, and 138 schools with an average 39 students in Hong Kong. Depending on the operationalization, the actual number of academically resilient students varied within countries from 137 to 5473 (for details see Appendix 1).

3.2 Measures

3.2.1 Disadvantaged student background: Adversity

PISA's composite SES index ESCS and three indicators of the SES subdimensions were used to examine how different conceptualizations of adversity affected the classification of students as academically resilient.

Economic, social, and cultural status The economic, social, and cultural status (ESCS) index is built on three components reflecting cultural and economic capital, thus two out of three of Bourdieu's dimensions of SES: parental education, parental occupation, and home possessions including books at home (OECD 2017). Social capital in the sense of Bourdieu (1986) is not covered by the ESCS.

The composite index is routinely used in resilience studies in the ILSA context; therefore, it was used in this study as well. The fixed threshold for ESCS was set to -0.68 across countries and reflected the bottom 1/3 of students internationally. The relative within-country thresholds were set to 0.25 , -1.69 , and -1.03 for Norway, Peru, and Hong Kong respectively and reflected the bottom 1/3 of the students nationally. Since the ESCS index was built on three standardized components via principal component analysis, some students may have the same ESCS score. For example, 25 students in Norway had the same ESCS score of 0.25 that represented the threshold. In this case, as many students were randomly selected out of the group of students with the same score as needed to end up with a group size of exactly 1/3. In the case of Norway, this meant to select 6 students (for details see Appendix 2).

Wealth As an index of students' economic capital, PISA provides an IRT scaled index called WEALTH, based on the number of material possessions. It includes 3 country-specific items and 9 items not directly related to educational support at home such as "Rooms with a bath or shower". The fixed threshold for WEALTH was set to -0.72 across countries reflecting the bottom 1/3 of all students internationally. The relative within-country thresholds were set to 0.26 , -2.52 , and -1.12 for Norway, Peru, and Hong Kong respectively.

Books This study used the variable "Number of books at home" (BOOKS) as an indicator of students' cultural capital. It consisted of 6 categories from 0 to 10 books to more than 500 books. The fixed threshold across countries was set to the second category (11–25 books), which included together with the first category a bit more than 1/3 of all students internationally. The relative within-country thresholds (bottom 1/3 of the students in each country) were set to the third category (26–100 books) for Norway, and to the second for both Peru and Hong Kong. In each country, several cases were in the category that reflected the threshold. To keep the number of disadvantaged students to the bottom 1/3 as intended, random cases were selected.

Parents' emotional support scale (EMOSUPS) Assessing students' social capital in ILSAs is challenging. Because it is typically represented by the relationship among family members that enhance the transmission of other resources (Bourdieu 1986), such as the family structure, parent-children discussion, parents' expectations and

aspirations of children, parental education style, or intergeneration closure (Dika and Singh 2002). Since neither Norway nor Peru participated in the parent's questionnaire, which included several social capital indicators, our study used the parents' emotional support scale (EMOSUPS) from the student questionnaire as a proxy. The IRT scaled index was based on four statements such as whether students perceived their parents as supportive when they faced difficulties at school. The fixed threshold across countries was set to -0.43 , and the relative within country-thresholds were -0.43 , -0.89 , -0.89 for Norway, Peru, and Hong Kong respectively. Random cases were selected if there were cases with the same score at the threshold.

To investigate whether classifications of resilience could be sensitive to the choice of the background indicator, we estimated Pearson correlations between them. The results indicated significant positive but imperfect associations among the four indicators (ESCS, WEALTH, BOOKS, and EMOSUPS). Basically, the composite index ESCS had strong or moderate associations with the indicators of economic (WEALTH) and cultural (BOOKS) capital, but only small associations with the indicator of social capital (EMOSUPS). Therefore, it is likely that the classification of disadvantaged students varies when different background indicators are applied to define student background.

3.2.2 Strong educational outcomes: positive adaptation

The outcome measure in this study is represented by the science score from PISA 2015. We used all 10 plausible values provided and combined the results of the ten separate analyses using Rubin's (1987) rules.

The fixed threshold was set at the mean of the PISA 2015 cycle of 466 points across countries. Disadvantaged students who scored higher than 466 were considered as resilient. The relative within-country thresholds, which represented the top 1/3 of the students within each education system, were set to 427.22 points for Peru, 542.61 points for Norway, and 561.13 points for Hong Kong.

3.2.3 Validity measures

Two student characteristics, gender and language spoken at home, were used to compare compositions of academically resilient students under different conceptualizations. Two external constructs, sense of belonging and absence from school, were adopted to examine the differences in strength between academic resilience and external constructs supposed to assess the same underlying idea across different definitions.

Gender Female students were coded as 1, male as 2. Gender is balanced in our sample. Further, 49.6% of the students were female in both Norway and Peru, and 49.5% were female in Hong Kong.

Language Students who usually used the same language as the language of the assessment were labeled as 1, and students who usually used another language were labeled as 2. About 8.7% and 7.2% of the students in Norway and Peru usually spoke another language at home; the proportion was smaller in Hong Kong, about 3.5%.

Sense of belonging Students were asked to rate six statements about their sense of belonging to school on a four-point Likert scale. We recoded the items so that higher scores referred to higher sense of belonging and built a latent variable BEL.

Absence from school PISA 2015 had three items to assess how often students skipped full school days, single classes, or arrived late at school. The four response categories ranged from “never” to “five or more times” with higher scores referring to higher absence from school. The three items were used to build the latent variable ABS.

Since BEL and ABS were both built on observed items assessed by a four-point Likert scale, we adopted Desa's (2014) suggestion and treated these items as categorical. A multiple-group confirmatory factor analysis (MG-CFA) was carried out to test measurement invariance of these constructs across Norway, Peru, and Hong Kong. This method usually relies on two-group comparison and is applied to relatively small sample sizes (Rutkowski and Svetina 2014). Therefore, comparisons across more than two groups with large sample sizes add complexity (Svetina et al. 2020). This study followed Svetina and Rutkowski's suggestion (2017) and adjusted the typical criteria to consider the changes in CFI greater than or equal to -0.004 and changes in RMSEA less than or equal to 0.050 for evaluating metric invariance, and changes in CFI greater than or equal to -0.004 and changes in RMSEA less than or equal to 0.010 for evaluating scalar invariance (Svetina et al. 2020). Analyses were conducted for BEL and ABS separately, metric invariance, and partial scalar invariance were established for both constructs (for details see Appendix 4).

3.3 Data analysis

The combination of four indicators of student background (1 composite SES index and 3 subdimensions of human capital) with two thresholds each time (1 fixed across countries and 1 relative within-country) led to eight operationalizations of adversity. In combination with the two versions of the indicator of positive adaption (science score with a fixed cross-country or a relative within-country threshold), we ended up with 16 different conceptualizations of academic resilience. The composition (gender and language spoke at home) of disadvantaged students were compared with the composition of the whole sample.

A logistic regression was fitted to estimate the likelihood of being resilient predicted by BEL and ABS. Data were handled and prepared in R (Version 3.6; R Core Team 2019), and analyses were conducted in Mplus (Version 8.4; Muthén and Muthén 2017). Missing data were handled by the default Full Information Maximum Likelihood (FIML) method in Mplus.

Students are nested in classes in our data set, but the nested data structure is not central to the research questions because all variables in question are located on the lowest level. Therefore, single-level models were estimated, which require fewer distributional assumptions and use a more parsimonious model approach (Stapleton et al. 2016). In the single-level logistic regression, student weights were used to make valid estimates and inferences of the population. Cluster characteristics due to the non-independence of samples (Stapleton et al. 2016) were taken into consideration by applying a sandwich estimator (type = complex in Mplus) and a robust weighted least

squares estimator (WLSMV) so that standard errors and fit statistics were calculated properly (Asparouhov and Muthén 2006) (Fig. 1).

Ten plausible values combined with one background indicator led to 10 results, thus the resilient status of a student can be different in these 10 results. While these were combined in the estimation of the regression coefficients, results based on the first plausible value are presented to provide descriptive information about how the compositions of students change under different conceptualizations of academic resilience.

Two types of significant tests were applied. Firstly, a two sample *t* test was used to test the difference between proportions of disadvantaged students who are resilient when fixed and relative background thresholds were applied, or when fixed and relative performance thresholds were applied. The same type of *t* test was used to test differences between proportions of female and male students classified as academically resilient students or between students who speak the language of the test or another language. Secondly, a Wald chi-square test was used to examine whether coefficients of BLE and ABS differ across conceptualizations.

4 Results

4.1 How do different conceptualizations of academic resilience affect how much and which students are classified as academically resilient?

4.1.1 First step: defining disadvantaged students (adversity)

By implication, when relative thresholds for students' background indicators were applied, one out of three students in each country were classified as a disadvantaged student. Therefore, proportions of disadvantaged students were around 33.33% across conceptualizations using a relative background threshold. When the same fixed thresholds for

Fig. 1 Single-level logistic regression model. *RES* resilient, *BEL* sense of belonging, *ABS* absent from school, st034q01ta to st034q06ta are six observed items for sense of belonging, st062q01ta to st062q03ta are three observed items for absent from school.

background indicators were used across countries, results varied substantially (see Table 2). For example, if the fixed threshold of ESCS (-0.68) was used to define student background, almost two-thirds of the students in Peru (63.94%) and half of the students in Hong Kong (46.76%) were considered as disadvantaged, but considerably fewer students in Norway (7.18%). Similar shifts were observed for the indicators of economic capital WEALTH and cultural capital BOOKS although on different levels.

In contrast, applying the background indicator of social capital, parents' emotional support (EMOSUPS) revealed a different pattern. While there were roughly one out of three students classified as disadvantaged in Norway (32.86%) and Peru (40.14%), the proportion was now highest in Hong Kong with almost two out of three students classified as disadvantaged (63.30%).

4.1.2 Second step: defining well-performing students (positive adaptation)

When relative within-country thresholds for science achievement were applied to define positive adaption, one out of three students was classified as well-performing in Norway, Peru, and Hong Kong. However, results varied substantially when the fixed threshold—set to the international PISA 2015 mean of 466 score points—was applied. Whereas in Norway almost two out of three students (63.54%) and in Hong Kong even more than three out of four students (78.17%) scored above the PISA 2015 mean, it was less than one out of five students in Peru (18.72%).

4.1.3 Third step: conceptualizing academically resilient students by combining background and performance definitions

There are in principle four possibilities to conceptualize academic resilience given what we have described above: combining a fixed or relative background threshold with the fixed performance threshold, or combining a fixed or relative background threshold with the relative performance threshold. We will systematically look at the results of these four approaches, firstly for the composite ESCS index and thereafter for the three subdimensions of human capital.

Combining the ESCS background thresholds with the fixed performance threshold When the fixed performance threshold of 466 points was combined with the ESCS indicator, about 70% of the disadvantaged students in Hong Kong and about half of the disadvantaged students in Norway were classified as being resilient, no

Table 2 Proportion of students classified as disadvantaged with fixed background thresholds

	Norway	Peru	Hong Kong
ESCS	7.18%	63.94%	46.76%
WEALTH	2.51%	78.51%	55.66%
BOOKS	20.80%	65.77%	36.41%
EMOSUPS	32.86%	40.14%	63.30%

ESCS index of economic, social, and cultural status; *EMOSUPS* index of parents' emotional support

matter whether the fixed or the relative background thresholds were applied. Less than 10% of the disadvantaged students in Peru were classified as being academically resilient in both cases.

Given the low share of Norwegian students classified as disadvantaged with the fixed ESCS threshold applied across countries, the overall share of all students classified as academically resilient was very low (2.99%). The same low proportion of academically resilient students applied to Peru, but here because of the low share of students scoring above the PISA 2015 mean. In contrast, either one out of three or four Hong Kong students was classified as academically resilient (Table 3).

Combining the ESCS background thresholds with the relative performance thresholds When relative performance thresholds were applied, one out of three students in each country was defined as well performing. The proportions of disadvantaged students classified as academically resilient were more similar than in the cases described above with fixed performance thresholds across countries. The share varied only between 13.17% in Peru and 28.93% in Hong Kong, no matter whether the fixed or the relative background threshold was applied (see Table 4). Overall, this meant that the proportion of disadvantaged students classified as academically resilient went up in Peru and down in Norway and Hong Kong, when relative performance thresholds were applied instead of the same fixed performance threshold.

The same pattern occurred if we look at the proportion of all students classified as academically resilient. For example, with a fixed threshold of ESCS, the proportion of all students classified as academically resilient in Norway went down to 0.97% but went up in Peru to 14.55%.

Applying the subdimensions of human capital to define adversity The ESCS is a composite index that includes economic, cultural, and social capital indicators of student background at the same time. If we disentangle the conceptualization of adversity by applying the indicators of the three subdimensions separately, the data revealed similarities but also substantial differences. Depending on the type of thresholds applied either to background or to performance, the proportion of students classified as academically resilient varied.

In case of the same fixed performance threshold across countries, the pattern was similar. The proportion of disadvantaged students classified as academically resilient was highest in Hong Kong (between two thirds and three quarters of the disadvantaged students), followed

Table 3 Proportion of students classified as academically resilient when the fixed performance threshold was applied with the composite ESCS index

	Proportion of disadvantaged students classified as resilient			Proportion of all students classified as resilient		
	Norway	Peru	Hong Kong	Norway	Peru	Hong Kong
Fixed ESCS	41.58%	9.85%	73.18%	2.99%	6.30%	34.22%
Relative ESCS	50.06%	4.71%	70.71%	16.16%	1.56%	23.07%

ESCS index of economic, social, and cultural status

Table 4 Proportion of students classified as academically resilient when relative performance threshold were applied with the composite ESCS index

	Proportion of disadvantaged students classified as resilient			Proportion of all students classified as resilient		
	Norway	Peru	Hong Kong	Norway	Peru	Hong Kong
Fixed ESCS	13.52%	22.75%	28.93%	0.97%	14.55%	13.53%
Relative ESCS	20.89%	13.17%	26.66%	6.75%	4.37%	8.70%

ESCS index of economic, social, and cultural status

by Norway (between about 40% and 60% of the disadvantaged students) and lowest in Peru (between 5% and 17% of the disadvantaged students)—no matter whether the composite ESCS index or its subdimensions WEALTH, BOOKS, or EMOSUPS and no matter whether the fixed or the relative background thresholds were applied.

Although the pattern was similar, differences in the actual group size were visible resulting from variation in the proportion of students classified as disadvantaged (see Table 2). The indicator of economic (WEALTH) and in particular of social capital (EMOSUPS) led to higher shares of disadvantaged students classified as academically resilient than the indicator of cultural capital (BOOKS), no matter which type of background threshold was applied. This was particularly visible in Peru, where the proportion was up to twice or even three times as high as if an indicator of another subdimension or the composite ESCS index had been used (for details see Appendix 5).

The differences between applying the composite ESCS index or one of the subdimensions of human capital as indicators of adversity were even more pronounced when relative performance thresholds were used to define positive adaptation. The proportion of disadvantaged students classified as academically resilient was no longer highest in Hong Kong or lowest in Peru. For example, when relative performance thresholds were applied together with BOOKS as the indicator of cultural capital or EMOSUPS as the indicator of social capital, the proportion of disadvantaged students classified as academically resilient was lowest in Norway, no matter whether a fixed or a relative background threshold was used.

Similar to applying the fixed performance threshold, using WEALTH or EMOSUPS as capital indicators together with the relative performance threshold resulted usually in larger proportions of students classified as academically resilient than BOOKS (with one exception in Peru), regardless of whether a fixed or relative background threshold was used. It was particularly in Peru where the application of the indicator for social capital (EMOSUPS) increased the proportion of disadvantaged students classified as academically resilient (for details see Appendix 6).

4.2 Which students are classified as academically resilient in the different conceptualizations?

The proportions of students classified as academically resilient varied a lot by gender and language depending on the conceptualization and the country. Using the ESCS

index as an indicator of adversity to define student background, there were between 31 female students (i.e., 0.44% of all students) in Peru and 905 female students (i.e., 16.89%) in Hong Kong classified as academically resilient, while the proportion of males varied between 0.37% (i.e., 20) in Norway and 17.34% (929) in Hong Kong (see Table 5). There were significantly fewer female than male students classified as academically resilient in Peru no matter which threshold was applied to define adversity or positive adaptation. If a relative outcome threshold was used, the same applied to Hong Kong. In contrast, Norway's proportions of female and male students classified as academically resilient were generally more balanced.

With respect to the language spoken at home in relation to the test language, there were between 2 students with a different language (i.e., 0.04% of all students) in Hong Kong and 97 students with a different language (i.e., 1.78%) in Norway classified as academically resilient, while the proportion of students with the same language classified as academically resilient varied between 0.88% (i.e., 48) in Norway and 33.74% (1808) in Hong Kong (see Table 6). In all three countries, there were significantly more students classified as academically resilient who spoke the test language at home, no matter which threshold was used to define adversity or positive adaptation.

Since the conceptualization of academic resilience included two criteria, namely student background with respect to adversity and student outcome with respect to positive adaptation, it was necessary to look at the criteria step-wise to be able to interpret the numbers presented above. We examined therefore the classification of students by gender and language firstly with respect to who was classified as disadvantaged, then at those who were classified as having high outcomes before we finally interpreted the combination in terms of academic resilience. As an overview, we started with applying ESCS as an indicator of student background, but we looked also at whether there were differences with respect to the economic, cultural, and social subdimensions of human capital.

Classification of academic resilience by gender There was a balanced gender distribution in Norway with respect to the classification of students as disadvantaged if the ESCS index was used as an indicator of adversity (see Appendix 9a). This result was

Table 5 Number of academically resilient students and percentage of all students by gender in each country

Back	Out	Norway				Peru				Hong Kong			
		Female		Male		Female		Male		Female		Male	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Fix	Fix	94	1.72	69	1.26	159	2.28*	280	4.02	905	16.89	929	17.34
	Rel	33	0.60	20	0.37	427	6.13*	587	8.42	324	6.05*	401	7.48
Rel	Fix	463	8.49	419	7.68	31	0.44*	78	1.12	608	11.35	628	11.72
	Rel	190	3.48	178	3.26	120	1.72*	185	2.65	210	3.92*	256	4.78

Back student background (ESCS), *Out* student outcome (science achievement), *Fix* same fixed threshold across countries, *Rel* within-country threshold, * = proportion of female students significantly different from the proportion of males within the same operationalization, $p < 0.05$

Table 6 Number of academically resilient students and percentage of all students by language in each country

Back	Out	Norway				Peru				Hong Kong			
		Diff L		Same L		Diff L		Same L		Diff L		Same L	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Fix	Fix	31	0.57*	132	2.42	7	0.10*	432	6.20	26	0.49*	1808	33.74
	Rel	5	0.09*	48	0.88	21	0.30*	993	14.24	2	0.04*	723	13.49
Rel	Fix	97	1.78*	785	14.39	4	0.06*	105	1.51	21	0.39*	1215	22.67
	Rel	33	0.60*	335	6.14	13	0.19*	292	4.19	2	0.04*	464	8.66

Back student background (ESCS), *Out* student outcome (science achievement), *Fix* same fixed threshold across countries, *Rel* within-country threshold, *L* language of the test spoken at home (different vs. same); * = the proportion of students who spoke another language at home was significantly different from the proportion of students who spoke test language at home in the same operationalization, $p < 0.05$

independent of the type of threshold applied. The proportions of female and male students in the top 1/3 Norwegian performers, i.e., in the group of those showing positive adaptation, were also evenly distributed—at least as long as the fixed performance threshold was used (see Appendix 11a). Applying the stricter relative performance threshold, there were significantly fewer females than males belonging to the top 1/3. As documented in Table 5, these performance differences were not large enough to affect the final gender distribution of academically resilient students, but they led to a significantly lower mean performance of the female students classified as academically resilient compared to males when the stricter relative within-country threshold in defining adversity was used (see Appendix 10a).

With respect to variation by subdimension of human capital (see Appendix 9a), the proportion of female students in Norway classified as disadvantaged was significantly lower than the proportion of males when BOOKS or Parents' emotional supports (EMOSUPS) were applied as indicators of adversity, independently of the type of threshold used. In contrast, significantly more female than male students were classified as disadvantaged, when the indicator WEALTH was applied together with the relative but not with the fixed threshold. The relative within-country threshold was much higher than applying the same fixed cross-country threshold.

Similar to Norway, if the ESCS was used as an indicator of adversity, there was a balanced gender distribution among disadvantaged students in Peru, independently of the type of threshold for adversity (see Appendix 9a). In Peru, the proportion of female students in the top 1/3 performers was significantly lower than of males, no matter which threshold indicating positive adaptation was applied (see Appendix 11a). As documented in Table 5, these performance differences led in turn to a significantly lower proportion of female than male students classified as academically resilient. Furthermore, the proportion of female students classified as academically resilient was significantly lower than the proportion of female students classified as disadvantaged in all operationalizations. In addition, the mean performance of academically resilient female students was often lower compared to male students, in particular when the more lenient relative within-country performance threshold was used (see Appendix 10a). Variation by subdimension of human capital was limited (see Appendix 9a).

Similar to Norway and Peru, there was no systematic difference in the gender distribution with respect to students' classification as disadvantaged using the ESCS index in Hong Kong (see Appendix 9a). The proportions of female and male students in the top 1/3 performers were evenly distributed if the lenient fixed performance threshold was used (see Appendix 11a). However, similar to Norway, applying the stricter relative performance threshold, there were significantly fewer females than males belonging to the top 1/3 (see Appendix 11a). Furthermore, the mean performance of academically resilient female students was often lower compared to male students (see Appendix 10a). With respect to the subdimensions of human capital, the same pattern was visible for Hong Kong as it had been documented for Norway when it came to BOOKS and EMOSUPS (see Appendix 9a). The proportion of female students classified as disadvantaged was significantly lower in these cases than of males, independently of the type of threshold used.

Classification of academic resilience by language Students who spoke a language at home different from the test language were significantly overrepresented in the group of students classified as disadvantaged with the ESCS index as an indicator of adversity in Norway (see Appendix 9b). This result was independent of the type of threshold applied. In contrast, they were significantly underrepresented in the top 1/3 Norwegian performers, i.e., in the group of those showing positive adaptation, again independently of the threshold used (see Appendix 11b). These differences strongly affected the distribution of academically resilient students by language (see Table 6). With respect to variation by subdimension of human capital (see Appendix 9b), when parents' emotional support (EMOSUPS) was applied as an indicator of adversity, independently of the type of threshold used, students who spoke another language at home were still significantly underrepresented but less dramatically as with the ESCS index.

The pattern was similar in Peru. Students who spoke a language at home different from the test language were significantly overrepresented in the group of students classified as disadvantaged no matter which threshold was used (see Appendix 9b). In contrast, they were significantly underrepresented in the top 1/3 performers, again independently of the threshold (see Appendix 11b). These differences strongly affected the distribution of academically resilient students by language (see Appendix 6). With respect to variation by subdimension of human capital (see Appendix 9b), the data revealed that using the WEALTH indicator together with the relative within-country threshold led to a stronger overrepresentation than with the other subdimensions.

The patterns were partly different in Hong Kong compared to Norway and Peru. There was no systematic effect of language on the classification as disadvantaged if the ESCS index was used (see Appendix 9b). However, students who spoke a different language at home were significantly underrepresented in the top 1/3 performers, independently of the threshold applied (see Appendix 11b). These performance differences strongly affected the distribution of academically resilient students by language (see Table 6). In regard to variation by subdimension of human capital (see Appendix 9b), the data revealed that using the BOOK indicator no matter which threshold was applied and using the relative WEALTH indicator led to a stronger overrepresentation of students with a different language than with the other subdimensions.

4.3 How are different conceptualizations associated with external variables?

In this section, we present the results of our study on concurrent validity including two external variables supposed to assess the same underlying idea. We present the relation between a classification of students as academically resilient and, firstly, their sense of belonging to a school and, secondly, their absence of school in terms of odds ratios based on standardized results (for details about model fit, estimates and 95% CI see Appendix 3).

Sense of belonging as the validity criterion Using the composite ESCS background indicator, sense of belonging (BEL) was statistically significantly and positively associated with academic resilience in all conceptualizations in Peru as hypothesized, but not in Norway or Hong Kong (see Table 7). These results indicate that students who said that they felt a higher sense of belonging to their school had a 30 to 40% higher chance to be classified as academically resilient in Peru. Unexpectedly, this did not apply to Norway or Hong Kong where no statistically significant effects of students' sense of belonging to their school on the classification as academically resilient was found.

With respect to the research question of differential results depending on the operationalization applied to academic resilience, there were no statistically significant differences for coefficients of BEL between fixed and relative performance thresholds or between fixed and relative background thresholds in case of using ESCS as an indicator of adversity. The same applied to using the number of books (BOOKS) or parents' emotional support (EMOSUPS) in all educational systems (see Appendix 7). However, there were statistically significant differences for coefficients of BEL between fixed and relative background thresholds in Peru when WEALTH was used.

Absence from school as the validity criterion Using the composite ESCS index as the background indicator, absence from school (ABS) was statistically significant and negatively associated with academic resilience in both Norway and Hong Kong as hypothesized (see Table 8). However in Peru, ABS was unexpectedly not significant in most operationalizations, and no noticeable pattern was found. A higher ABS score refers to higher frequency of absence from school; the results indicate that in Norway and Hong Kong, students who were more frequently absent from school had a 25 to 30% lower chance to be academically resilient.

Table 7 Odds ratio of effects of sense of belonging on the classification as academically resilient when ESCS was applied

		Norway ESCS		Peru ESCS		Hong Kong ESCS	
		Fixed	Relative	Fixed	Relative	Fixed	Relative
Performance	Fixed	1.05	1.00	1.30*	1.39*	0.95	0.97
	Relative	1.02	0.99	1.35*	1.40*	0.94	0.92

* $p < 0.05$, ESCS index of economic, social, and cultural status

Table 8 Odds ratios of effects of absence from school on the classification as academically resilient when ESCS was applied

		Norway		Peru		Hong Kong	
		ESCS		ESCS		ESCS	
		Fixed	Relative	Fixed	Relative	Relative	Relative
Performance	Fixed	0.72*	0.70*	0.94	1.00	0.69*	0.72*
	Relative	0.75*	0.71*	0.98	1.04	0.68*	0.69*

* $p < 0.05$, *ESCS* index of economic, social, and cultural status

With respect to the research question of differential results depending on the operationalization of academic resilience, there were no statistically significant differences for coefficients of ABS between fixed and relative performance thresholds or between fixed and relative background thresholds in Norway and Hong Kong, neither in case of ESCS nor in case of one of the subdimensions of human capital. This applied to Peru when the number of books (BOOKS) and parents' emotional support (EMOSUPS) were applied, but not for ESCS and WEALTH (see Appendix 8).

5 Discussion

This study examined the conceptualization and operationalization of academic resilience by firstly identifying adversity with indicators of student background, then by identifying positive adaptation with a performance indicator, and finally by combining these two criteria. In this process, many decisions had to be made: which indicator should be selected (e.g., composite or specific indicators in case of adversity), which type of threshold should be applied (the same fixed one across countries or relative within-country ones), which level should be set on each threshold (strict or lenient ones). The key finding of the present study was that the way how academic resilience was defined mattered. Different decisions resulted in different proportions of students who were classified as academically resilient and different compositions of the group of resilient students.

Our analyses revealed that fixed thresholds were frequently not well suited to identify academically resilient students across diverse countries. This applied to the background indicators as well as for the performance indicator, as the distributions of SES and performance measures varied considerably across economically developing and developed economies. For example, given the low share of Norwegian students classified as disadvantaged with the same fixed background threshold applied across countries, the overall share of all students classified as academically resilient was very low. The low proportion of academically resilient students was also found in Peru. However in this case, it was because of the low share of students scoring above the fixed performance threshold. In contrast, in Hong Kong, either one out of three or four students was classified as academically resilient in these cases. When a fixed threshold

is used to define positive adaptation, the share of academically resilient students is heavily influenced by the level of economic development. As it was done for example by Sandoval-Hernández and Bialowolski (2016) or Erberer et al. (2015), a large share of academically resilient students does not necessarily equal to a better educational system as it was concluded in some reports (e.g., OECD 2011).

Generally, research on academic resilience is working with small sample sizes. This means we deal with strongly selected groups, which is not only a measurement challenge but raises policy questions as well. To what extent is it meaningful to implement activities in such a case? This challenge is increased by applying the same threshold across countries. For example, when a fixed ESCS threshold was combined with a relative performance threshold, only 53 Norwegian students were classified as academically resilient. Considering the whole sample size of Norway (5456), information about so few students is questionable if it shall be used to derive general measures beyond support of a specific group, which often is the conclusion in academic resilience papers.

SES refers to an individual or a family's position in a hierarchy according to access to wealth, power, and social status, and it usually includes parents' occupation, parents' education, and home income (Watermann et al. 2016). However, there were several concerns about these components. Researchers tended to use proxy to measure home income, but as mentioned before, owning a car means very different across countries. Measures about parents' occupation and education were usually collected from students' questionnaire, but students' responses may suffer from some degree of error, and the discordance varies across countries (Rutkowski and Rutkowski 2013). Therefore, comparisons of SES across country, for example, using a fixed SES threshold across countries, tend to raise measurement issues.

When relative within-country performance thresholds were applied to indicate positive adaptation, measurement problems with the student background indicators are no longer relevant. Furthermore, the proportions of disadvantaged students classified as academically resilient were more similar across countries than with the fixed performance threshold, because the proportions went up in Peru and down in Norway and Hong Kong. However, it seems worth to emphasize that the relative performance thresholds are hardly comparable in absolute terms, because they differed by 134 points (Peru 417 and Hong Kong 561), which corresponds to almost one and a half standard deviation on the PISA scale.

Based on Bourdieu's capital theory (1986), we adopted the three subdimensions of students' human capital separately to disentangle the effects of the composite ESCS conceptualization on academic resilience. The data revealed many similarities in the results but also substantial differences. The indicator of social capital led to higher shares of disadvantaged students classified as academically resilient than applying the composite index, and this particularly in Peru. This result may indicate a cultural difference between Peru and the other two systems, with social capital being educationally more relevant in Peru than in Norway and Hong Kong and thus more often revealed in the student survey in the first than in the latter systems. It would fit to results from cultural psychology where Peru often is characterized as a so-called collectivist country where social relations are more important than in so-called individualist countries such as Norway where people act rather independently (Hofstede and Peterson 2003).

The indicator of cultural capital resulted in all conceptualizations in smaller proportions of disadvantaged students classified as academically resilient than the composite ESCS index. This result may indicate a large spread in cultural capital than in the other types of human capital between disadvantaged and advantaged children. It seems to be harder for disadvantaged students to accomplish an amount of cultural capital that equals economic or social capital. The causal mechanism here is unknown though and should be examined in further research. It is worthwhile to point out that none of these differences resulting from applying one of the three subdimensions of human capital showed up when the composite ESCS index was used. The country differences in classifying disadvantaged students as academically resilient were often no longer significant. It seems as if advantages and disadvantages of the different approaches were balanced out in this case.

Concerning the composition of the group of students classified as academically resilient, the proportions varied by gender and language depending on the conceptualization and the country. This result points to the gender- and language-specific sensitivity of conceptualizations of academic resilience, and this in turn may reflect societal characteristics. It is therefore essential which operationalization of academic resilience is chosen in a study. The conclusions may be completely different. In many cases, it mattered for the composition of the group of academic resilient students whether the same fixed cross-country threshold was applied or a more or less lenient or strict relative within-country threshold. For example, applying the stricter relative performance threshold in Norway and Hong Kong led to significantly fewer females than males belonging to the top 1/3, which may be related to the general discussion about gender balance at the top of the performance distribution (Bergold et al. 2017). In Peru, the proportion of females was lower at the top no matter which threshold was applied.

It mattered also which subdimension of human capital was applied in several cases. The data revealed for example that both in Norway and Hong Kong the proportion of female students classified as disadvantaged was significantly lower than of males when the indicators of cultural or social capital were applied as indicators of adversity than the indicator of economic capital. Similarly, in two out of three countries, students who spoke a language at home different from the test language were less strongly under-represented when the social capital indicator was applied, but more strongly when the economic capital indicator was applied. Educational inequalities between male and female students in student achievement, for example in terms of grades in mathematics or reading (Voyer and Voyer 2014), are well known for many countries and most researchers are aware of them. The same applies to educational inequalities depending on the language background of students (OECD 2018). Our study shows that it is important to pay attention to similar inequalities when it comes to background indicators as part of the definition of academic resilience as well.

When the same background indicator was applied with different performance thresholds, proportions of female or male students classified as academically resilient or students who spoke another language at home varied as well. If a fixed performance threshold was adopted, we were likely to underestimate the inequality in high-performing systems but overestimate the inequality in low-performing systems, thus exaggerate the inequality differences between high- and low-achieving systems. For language-related analysis, the data revealed that students who spoke another language at home were overrepresented in disadvantaged students but underrepresented in resilient students.

Finally, the results of our validity study revealed that the likelihood of being classified as resilient most often did not change significantly when different background thresholds or performance thresholds were applied. This result indicates some consistency in their definitions. At least the average size of relations to external constructs may not be affected too heavily.

6 Limitations

Before we turn to conclusions, it is necessary to point out some limitations of this study. Regarding indicators used in this study, student's performance in science was the only outcome indicator; analyses based on this domain may be not hold for other domains. Although we used four background indicators, ESCS as a composite SES index, WEALTH as an economic capital measure, BOOKS as a cultural capital measure, and EMOSUPS as a social capital index, all these indicators have limitations. For example, student's questionnaire of PISA 2015 had several items about social activities before and after school, including communications between parents and students, which would have been a good indicator of social capital. However, there was a large proportion of missing data (about 30%) from Peru. Therefore, this study adopted EMOSUPS as the only indicator to measure background from a social capital perspective. Finally, we have to deal with small groups in some cases. For example, when it comes to the research question which students were classified as academically resilient, the number of female students was as low as 15 students in Norway, and the number of students who spoke another language at home varied between 0 in Peru and 97 in Norway. The small numbers lead to large standard errors.

7 Conclusions

Proportions of resilient students are results of two criteria—a disadvantaged student background in terms of adversity and a high educational outcome in terms of positive adaptation. Our study shows how important it is to be careful with how to define both. How many and which students are classified as academically resilient is likely to be affected by the developmental state of a country, when a fixed threshold is applied to identify disadvantaged students. Thus, the pool of students who have a chance to be classified as academically resilient is predefined differently in each country. Countries that are highly developed economically may not have many disadvantaged students from a global perspective, which means that they by definition can have only very few academically resilient students. Considering the country-specific characteristics of background indicators, it may therefore be more meaningful to adopt a relative background threshold to define adversity.

In contrast, a decision on the type of performance threshold should largely depend on the aim of a study. If one would like to look at academic resilience across countries, for example with a global labor market in mind, a fixed performance threshold seems to be most meaningful. If one would like to look at academic resilience within a country, for example with the national labor market in mind, a relative performance threshold provides most information. However, given that some studies adopted the proportions

of academically resilient students as an indicator for the quality and equity of an educational system (see, e.g., Agasisti et al. 2018), it is highly important also in this case to be aware of the changes in outcomes when different types of performance thresholds are applied. With a fixed performance threshold, we may conclude that Hong Kong has a higher level of quality and equity than Norway. If we replace the fixed performance thresholds with relative ones, proportions of resilient students in Norway and Hong Kong were no longer significantly different. Similarly, changes in the composition by gender and language have to be considered.

A conceptualization of academic resilience applying a relative background threshold and a fixed performance threshold was recommended by some studies (see, e.g., OECD 2011) to explore the presence of academic resilience across countries, or to study the consistency of individual and environmental characteristics associated with resilience. However, the shortcoming of this operationalization was also mentioned by OECD (2018). It may overestimate the amount of academic resilience in some countries while underestimate in some others. Since the share is closely related to students' average performance, higher-performing system tends to have bigger shares. However, it does not necessarily equal to a higher ability in helping disadvantaged students to exceed their predicted performance given their background, which is the essence of academic resilience. Although it may be too early to conclude that relative within-country thresholds work best also with respect to performance, these reflections point in any case to a strong need of more research that looks into the consequences of different conceptualizations of academic resilience. Researchers should feel encouraged to apply different approaches and to compare their results with respect to their robustness. Furthermore, they should not only pay attention to the overall proportion of students classified as academically resilient but also to their composition by gender or language.

Some conclusions can also be drawn regarding the choice of background indicators. The composite ESCS index as a multi-dimensional composite index seems to balance out to some extent different estimations happening when one of the subdimensions of human capital is applied. Whether this is appropriate is an open question and requires more research. The question here is, does an application of the subdimensions over- or underestimate the true size or do they indicate real differences.

Further research is thus needed in many respects. Research on academic resilience in education has deep roots in resilience studies carried out in psychology and sociology. However, there are differences we can learn from and which could move research on academic resilience forward. Firstly, unlike resilience studies in psychology, most studies in education are not longitudinal. It would be easier to identify causal mechanisms that increase academic resilience and may thereby contribute to overcoming educational inequality. Secondly, academic resilience studies have a methodological limitation in measuring adversity (Waxman et al. 2003), because disadvantaged students are treated as homogeneous groups despite possible variations in the degree to which their lives are actually affected by a risk (Luthar and Zelazo 2003). Our study has pointed to an approach to deal with this challenge by distinguishing between different types of human capital. Thirdly, most academic resilience studies in education treat non-cognitive skills as protective factors rather than including them as an indicator of positive adaptation. Acknowledging the relevance of, for example self-efficacy or educational aspiration as criterion for being academically resilience could change the discussion substantially.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2018). Academic resilience: What schools and countries do to help advantaged students succeed in PISA. OECD Education Working Papers, 167. <https://doi.org/10.1787/e22490ac-en>.
- Agasisti, T., & Longobardi, S. (2014). Inequality in education: Can Italian disadvantaged students close the gap? *Journal of Behavioral and Experimental Economics*, 52(C), 8–20. <https://doi.org/10.1016/j.socec.2014.05.002>.
- Agasisti, T., & Longobardi, S. (2017). Equality of educational opportunities, schools' characteristics and resilient students: An empirical study of EU-15 countries using OECD-PISA 2009 data. *An International and Interdisciplinary Journal for Quality-of-Life Measurement*, 134(3), 917–953. <https://doi.org/10.1007/s11205-016-1464-5>.
- Asparouhov, T., & Muthén, B. (2006). Comparison of estimation methods for complex survey data analysis. *Mplus web notes*.
- Aydiner, A., & Kalender, I. (2015). Student segments based on the factors related to sense of belonging across disadvantaged and resilient groups in PISA 2012. *Procedia - Social and Behavioral Sciences*, 174, 3299–3305. <https://doi.org/10.1016/j.sbspro.2015.01.997>.
- Bergold, S., Wendt, H., Kasper, D., & Steinmayr, R. (2017). Academic competencies: Their interrelatedness and gender differences at their high end. *The Journal of Educational Psychology*, 109(3), 439–449. <https://doi.org/10.1037/edu0000140>.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Westport, CT: Greenwood Press.
- Broer, M., Bai, Y., & Fonseca, F. (2019). *Socioeconomic inequality and educational Outcomes: Evidence from Twenty Years of TIMSS*. <https://doi.org/10.1007/978-3-030-11991-1>.
- Cheung, K. C. (2017). The effects of resilience in learning variables on mathematical literacy performance: A study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea. *Educational Psychology*, 37(8), 965–982. <https://doi.org/10.1080/01443410.2016.1194372>.
- Cheung, K. C., Sit, P. S., Soh, K. C., Jeong, M. K., & Mak, S. K. (2014). Predicting academic resilience with Reading engagement and demographic variables: Comparing Shanghai, Hong Kong, Korea, and Singapore from the PISA perspective. *The Asia-Pacific Education Researcher*, 23(4), 895–909. <https://doi.org/10.1007/s40299-013-0143-4>.
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: an introduction to tests and measurement (ninth edition. ed.)*. New York: McGraw-Hill Education.
- Coronado-Hijón, A. (2017). Academic resilience: A transcultural perspective. *Procedia - Social and Behavioral Sciences*, 237(C), 594–598. <https://doi.org/10.1016/j.sbspro.2017.02.013>.
- Desa, D. (2014). *Evaluating measurement invariance of TALIS 2013 complex scales*. <https://doi.org/10.1787/5f22kbbv1b7k-en>.
- Dika, S. L., & Singh, K. (2002). Applications of social Capital in Educational Literature: A critical synthesis. *Review of Educational Research*, 72(1), 31–60. <https://doi.org/10.3102/00346543072001031>.
- Economy, H. K. (2010). Hong Kong's Gini coefficient compared with other economies. Retrieved from https://www.hkeconomy.gov.hk/en/pdf/gini_comparison.pdf.
- Erberer, E., Stephens, M., Mamedova, S., Ferguson, S., & Kroeger, T. (2015). Socioeconomically disadvantaged students who are academically successful: Examining academic resilience cross-nationally. IEA Policy Brief series, 5. Retrieved from http://pub.iea.nl/fileadmin/user_upload/Policy_Briefs/IEA_Policy_Brief_Dec2013.pdf

- Fink, A. (2010). Survey research methods. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 152–160). Oxford: Elsevier.
- Frempong, G., Visser, M., Feza, N., Winnaar, L., & Nuamah, S. (2016). Resilient learners in schools serving poor communities. *Electronic Journal of Research in Educational Psychology*, 14(2), 352–367. <https://doi.org/10.14204/ejrep.39.15038>.
- García-Crespo, F. J., Galián, B., Fernández-Alonso, R., & Muñoz, J. (2019). Educational resilience in reading comprehension: Determinant factors in PIRLS-Europe, Resiliencia Educativa en Comprensión Lectora: Factores determinantes en PIRLS-Europa. *Revista de Educacion*, 2019(384), 71–96. <https://doi.org/10.4438/1988-592X-RE-2019-384-413>.
- Gonzalez, R., & Padilla, A. M. (1997). The academic resilience of Mexican American high school students. *Hispanic Journal of Behavioral Sciences*, 19(3), 301–317. <https://doi.org/10.1177/07399863970193004>.
- Hagenaars, A., & De Vos, K. (1988). The definition and measurement of poverty. *The Journal of Human Resources*, 23(2), 211–221. <https://doi.org/10.2307/145776>.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482. <https://doi.org/10.1086/504455>.
- Hofstede, G., & Peterson, M. (2003). Culture's consequences: Comparing values, behavior, institutions, and organizations across nations. In (Vol. 48, pp. 127–131).
- Kalaycıoğlu, D. B. (2015). The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Educational Sciences: Theory and Practice*, 15(5), 1391–1401. <https://doi.org/10.12738/estp.2015.5.2731>.
- Kane, M., Crooks, T., & Cohen, A. (2005). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. <https://doi.org/10.1111/j.1745-3992.1999.tb00010.x>.
- Karklina, I. (2012). Family and children educational outcomes: social resilience within economically deprived families in Latvia. Paper presented at the rural environment. Education. Personality. (REEP). Proceedings of the international scientific conference Latvia.
- Luthar, S. S., & Zelazo, L. B. (2003). Research on resilience: An integrative review. *Resilience and vulnerability: Adaptation in the context of childhood adversities*, 2, 510–549. <https://doi.org/10.1017/CBO9780511615788.023>.
- Martin, A. J., & Marsh, H. W. (2006). Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychology in the Schools*, 43(3), 267–281. <https://doi.org/10.1002/pits.20149>.
- Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology*, 19(3), 921–930. <https://doi.org/10.1017/S0954579407000442>.
- Meyer, H. D., Strietholt, R., & Epstein, D. Y. (2017). Three models of global education quality and the emerging democratic deficit in global education governance. In A. M. & L. G. (Eds.), *International Handbook of Teacher Quality and Policy*. New York: Routledge.
- Mullis, I. V., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2004). PIRLS 2006 assessment framework and specifications: Progress in international reading literacy study. TIMSS & PIRLS, International Study Center, Lynch School of Education, Manressa House, Boston College, 140 commonwealth street, Chestnut Hill, MA 02467.
- Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 assessment frameworks. In Retrieved from Boston College. TIMSS & PIRLS International Study: Center website <http://timssandpirls.bc.edu/timss2015/frameworks.html>.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- OECD. (2005). PISA 2003: Technical report. Retrieved <http://www.pisa.oecd.org/dataoecd/49/60/35188570.pdf>, from OECD.
- OECD. (2010). *PISA 2009 results: Overcoming social background equity in learning opportunities and outcomes (volume II)*. Paris: OECD Publishing.
- OECD. (2011). *Against the odds: Disadvantaged students who succeed in school*. Paris: OECD Publishing.
- OECD. (2017). PISA 2015 technical report. Retrieved <https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>, from OECD.
- OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. Paris: PISA, OECD Publishing.
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reimer, D., Sortekar, B., Oskarsson, M., Nilsen, T., Rasmusson, M., & Nissinen, K. (2018). *Northern lights on TIMSS and PISA 2018*. Copenhagen: Nordisk Ministerråd.
- Rutkowski, D., Rutkowski, L. (2013). Measuring Socioeconomic Background in PISA: One Size Might not Fit all. *Research in Comparative and International Education*, 8(3), 259–278

- Rutkowski, L. & Svetina, D. (2014). “Assessing the hypothesis of measurement invariance in the context of large-scale international surveys.” *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>.
- Sandoval-Hernández, A., & Bialowski, P. (2016). Factors and conditions promoting academic resilience: A TIMSS-based analysis of five Asian education systems. *Asia Pacific Education Review*, 17(3), 511–520. <https://doi.org/10.1007/s12564-016-9447-4>.
- Sandoval-Hernandez, A., & Cortés, D. (2012). Factors and conditions that promote academic resilience: A cross-country perspective. In *Paper presented at the international congress for school effectiveness improvement (ICSEI)*. Malmö: Sweden.
- Stapleton, L. M., McNeish, D. M., & Yang, J. S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, 51(3–4), 317–330. <https://doi.org/10.1080/00461520.2016.1207178>.
- Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology*, 48(7), 991–1008. <https://doi.org/10.1177/0022022117717028>.
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. <https://doi.org/10.1080/10705511.2019.1602776>.
- Tommaso, A., Francesco, A., Francesca, B., & Sergio, L. (2018). Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA: OECD publishing.
- Tudor, K. E., & Spray, C. M. (2017). Approaches to measuring academic resilience: A systematic review. *International Journal of Research Studies in Education*, 7(4). <https://doi.org/10.5861/ijrse.2017.1880>.
- UNDP. (2016). Human development report 2016: Human development for everyone. Retrieved from http://hdr.undp.org/sites/default/files/2016_human_development_report.pdf.
- Ungar, M. (2005). Handbook for working with children and youth: Pathways to resilience across cultures and contexts.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://doi.org/10.1037/a0036620>.
- Wang, Haertel, G., & Walgberg, H. (1994). Educational resilience in inner-city. In M. Wang & G. E. (Eds.), *Educational resilience in inner-city America: Challenges and prospects* (pp. 45–72). Hillsdale: Lawrence Erlbaum Associates.
- Watermann, R., Maaz, K., Bayer, S., & Roczen, N. (2016). Social background. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective* (pp. 117–145). Cham: Springer International Publishing.
- Waxman, H. C., Gray, J. P., & Padron, Y. N. (2003). Review of research on educational resilience. In *eScholarship*. California: University of.
- Werner, E. E. (2000). Protective factors and individual resilience. *Handbook of early childhood intervention*, 2, 115–132.
- Windle, G. (2011). What is resilience? A review and concept analysis. *Reviews in Clinical Gerontology*, 21(2), 152–169. <https://doi.org/10.1017/S0959259810000420>.
- Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes*, 9(1), 8–8. <https://doi.org/10.1186/1477-7525-9-8>.

Affiliations

Wangqiong Ye¹ · Rolf Strietholt² · Sigrid Blömeke¹

✉ Sigrid Blömeke
sigrid.blomeke@cemo.uio.no Wangqiong Ye
wangqiong.ye@cemo.uio.no

Rolf Strietholt
rolf.strietholt@tu-dortmund.de

¹ Centre for Educational Measurement, University of Oslo, Oslo, Norway

² Technische Universität Dortmund, Dortmund, Germany