

IEA Research_ and Development_

Participation and Accommodation Policies in IEA-Assessments: Patterns of Inclusion and Exclusion

Ana Maria Mejia-Rodriguez¹
Nurullah Eryilmaz¹
Olga Kunina-Habenicht²
Rolf Strietholt¹

¹International Association for the Evaluation of Educational Achievement
²TU Dortmund University

Preprint DOI: https://doi.org/10.58150/rd_4_6

Summary

International educational assessments are central to monitoring educational quality worldwide. However, little is known about which students are systematically excluded from these assessments and what accommodations are provided to those who participate. This study surveys national research coordinators of international assessments across 63 countries to examine how students with 17 different difficulties are included, excluded, and accommodated in international assessments. Findings reveal that students with severe physical, sensory, or multiple disabilities are consistently the most excluded, while those with learning or behavioral difficulties are more frequently included. Accommodations concentrate on timing and setting adjustments, whereas specialized supports remain rare. Cross-national variation is substantial, indicating that participation reflects national implementation decisions as much as student characteristics. The study highlights a key challenge: assessment results may reflect differences in participation and accommodation practices as much as differences in educational outcomes. Improving transparency, accessibility, and inclusive assessment design is therefore essential to ensure that IEA data more accurately represent all learners and support global monitoring of inclusive and equitable education.

This work was supported in whole or part by the IEA Research and Development (R&D) Funds, which aim to advance and improve the science and methodology of IEA studies, ensuring that they remain at the forefront of international large-scale assessments in education. Funding is drawn from an overhead on participation fees across IEA studies. This does not constitute endorsement of the contents, which reflects the views only of the authors. IEA cannot be held responsible for any use which may be made of the information contained therein.

© Stichting IEA Secretariaat Nederland 2026. All commercial rights are reserved. A non-exclusive and non-commercial license was granted to the fund recipient to use the intellectual property for future purposes.

This work is licensed under a Creative Commons Attribution - NonCommercial 4.0 International License, which permits noncommercial 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 license, and indicate if changes were made. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>.

Images or other third-party material in the work are included in the Creative Commons license unless indicated otherwise in a credit line to the material. If any material is not included in the work's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you must obtain permission directly from the copyright holder.

Participation and Accommodation Policies in IEA-Assessments:
Patterns of Inclusion and Exclusion

Ana Maria Mejia-Rodriguez¹, Nurullah Eryilmaz¹, Olga Kunina-Habenicht², & Rolf Strietholt¹

¹ International Association for the Evaluation of Educational Achievement

² TU Dortmund University

Abstract

International educational assessments are central to monitoring educational quality worldwide. However, little is known about which students are systematically excluded from these assessments and what accommodations are provided to those who participate. This study surveys national research coordinators of international assessments across 63 countries to examine how students with 17 different learning difficulties or support needs are included, excluded, and accommodated in international assessments. Findings reveal that students with severe physical, sensory, or multiple disabilities are consistently the most excluded, while those with learning or behavioral difficulties are more frequently included. Accommodations concentrate on timing and setting adjustments, whereas specialized supports remain rare. Cross-national variation is substantial, indicating that participation reflects national implementation decisions as much as student characteristics.

Introduction

Over the past three decades, inclusive education has become a firmly established global policy priority and a fundamental human right. The UNESCO Salamanca Statement (1994), endorsed by 92 governments, articulated the principle that schools should accommodate all children regardless of their physical, intellectual, social, emotional, or linguistic conditions (UNESCO, 1994). This normative commitment was later codified into binding international law through the United Nations Convention on the Rights of Persons with Disabilities (CRPD; United Nations, 2006). This convention requires countries to ensure that students with disabilities are not excluded from the general education system and that they are provided with reasonable accommodations (United Nations, 2006). The CRPD Committee's General Comment No. 4 further clarified that inclusive education entails systemic transformation across policy, practice, and culture to remove barriers to participation (Committee on the Rights of Persons with Disabilities, 2016; UNESCO, 2009, 2017).

These commitments were reinforced through the Incheon Declaration (UNESCO, 2015), which established a global plan for achieving Sustainable Development Goal 4 (SDG 4): ensuring inclusive and equitable quality education for all (United Nations, 2015). Despite these strong policy frameworks, implementation remains uneven. The UNESCO Global Education Monitoring Report (2020) found that only 17% of countries had legislative provisions supporting inclusive education. However, recent data suggest that legislation is shifting rapidly: the share of countries whose laws permit segregation has fallen from 25% to 9%, while the share supporting inclusion has risen to 29%, with the majority of countries (59%) opting for a combination of mainstream and separate settings (UNESCO, 2025). Despite this legislative progress, only 24% of countries have enacted inclusive education laws, while 37% rely on policies rather than binding legislation (UNESCO, 2025).

This gap between policy commitments and educational practice underscores the importance of examining not only access to schooling but also participation in key mechanisms used to evaluate educational systems. In particular, understanding whether students with disabilities and other learning or support needs are included in the evidence base that informs global education policy is critical for assessing progress toward SDG 4 (Mezzanotte, 2023). However, examining the participation of these students across countries presents important conceptual challenges, as definitions of disability, special

educational (SEN) needs, and support needs vary considerably across countries (Brussino, 2020; OECD, 2012), making direct comparisons difficult. Therefore, recent research has shifted focus from disability labels to specific difficulties and participation needs (Guez et al., 2024). Rather than asking how countries respond to particular diagnostic categories, this approach focuses on the participation challenges students may face in assessment. This perspective is consistent with the International Classification of Functioning, Disability and Health (ICF; WHO, 2001), which views disability not as a fixed individual characteristic but as the result of interactions between individuals and their environments.

Drawing on the approach proposed by Guez et al. (2024), in this project we develop a pragmatic framework consisting of 17 difficulties spanning sensory, physical, cognitive, linguistic, behavioral, and health-related domains. In this report, the term *difficulties* is used as an umbrella term to capture a broad range of characteristics or circumstances that may affect students' participation in international assessments and the types of support they may require. The framework combines impairment-related and functional categories with other participation-relevant difficulties, reflecting a needs-based perspective focused on participation rather than fixed diagnostic categories. While no framework can capture the full diversity of students' needs and circumstances, focusing on difficulties provides a practical and internationally comparable basis for examining participation across countries.

Inclusive Assessment: Participation, Validity, and Equity

A critical yet frequently underexamined dimension of inclusive education is whether students with different difficulties can meaningfully participate in assessments. Educational assessments serve as the primary instruments through which system performance is evaluated, inequalities are identified, and policy decisions are informed. When students with certain difficulties, SEN, or disabilities are systematically excluded from these assessments, the resulting evidence base becomes incomplete and potentially biased.

From a measurement perspective, inclusion is not only a policy goal but also a validity requirement. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on

Measurement in Education [NCME], 2014) identify fairness as a central component of validity, emphasizing the need to minimize construct-irrelevant barriers that may prevent examinees from demonstrating their true abilities (Messick, 1989). Accessible assessment conditions and appropriate accommodations are therefore essential for ensuring that test scores accurately reflect the intended constructs.

Similarly, policy frameworks mandate the participation of students with disabilities in large-scale assessments with appropriate accommodations (Individuals with Disabilities Education Improvement Act, 2004; National Center for Education Statistics, 1996). These principles reflect a broader understanding that inclusive assessment is inseparable from inclusive education (Quenemoen & Thurlow, 2019; Wagemaker, 2020). Despite these principles, the equitable participation of students with difficulties in assessments remains challenging across educational systems. Students with difficulties are often omitted from testing due to a perceived inability to participate, a lack of appropriate accommodations, or institutional practices that place them outside assessment sampling frames (Ketterlin-Geller, 2008). However, inclusion in an assessment does not by itself ensure equitable participation, as barriers related to assessment design, administration, or available supports may affect students' ability to demonstrate their knowledge and skills. As a result, assessment outcomes may reflect not only differences in educational performance but also differences in participation practices, raising concerns about validity and comparability.

Beyond measurement concerns, the systematic exclusion of students from assessments also raises fundamental human rights issues. When students with difficulties are omitted from the evidence base, they are effectively erased from educational and policy discourse (Schuelka, 2013). Research in the Norwegian context has shown that such practices can reinforce the perception that excluded students do not belong in mainstream educational settings (Aursand & Rutkowski, 2021). Exclusion from assessment is thus not only a technical challenge but also a normative one, with consequences for how educational systems recognize and respond to student diversity.

Participation and Representation in International Assessments

The tension between inclusion and assessment practices is also evident in international large-scale assessments. Studies such as IEA TIMSS, PIRLS, ICCS, and ICILS, as well as OECD PISA, play a central role in monitoring educational systems and informing global policy agendas (Klieme, 2012; Strietholt & Johansson, 2023). These assessments rely on complex sampling and implementation procedures that shape which students are ultimately included in the assessment data (LaRoche et al., 2020; Almaskut et al., 2023; Meinck, 2020).

While these assessments attempt to represent the full student population at certain age or grade levels, a fundamental question remains insufficiently addressed: who is actually represented in these assessments? Students facing certain participation difficulties, including those with functional disabilities, intellectual disabilities, or with limited proficiency in the language of assessment are frequently excluded from international assessments. These categories reflect groups commonly identified in the survey operations manuals and technical reports of these assessments, although the specific wording may vary slightly across studies. As with other assessments, such exclusions raise significant concerns for both equity and measurement validity. When identifiable subgroups are systematically omitted, national achievement estimates and responses to the background surveys may fail to represent the full student population, and differences observed across countries may partly reflect variation in participation practices rather than true differences in educational outcomes (Klieme, 2012; Schuelka, 2013).

Recent research has documented increasing exclusion rates and substantial variation across countries participating in international assessments (Atasever et al., 2024; Kjeldsen & Strietholt, 2024; Andersson & Sandgren Massih, 2023). Public debates have also highlighted concerns about the potential impact of exclusion practices on national performance indicators (Bjerril, 2024). These patterns suggest that exclusion is not merely a technical necessity but is also shaped by national interpretations of eligibility criteria and implementation practices (Cortes & Atasever, 2024; Meinck, 2020).

Assessment Accommodations as a Mechanism for Inclusion

Assessment accommodations represent a key mechanism for facilitating the participation of students with difficulties in standardized testing contexts. Accommodations are designed to reduce construct-irrelevant barriers by modifying aspects of test administration—such as timing, presentation format, response mode, or testing environment—without altering the construct being measured (Fuchs & Fuchs, 1999; Lovett & Lewandowski, 2015). When appropriately implemented, accommodations can enhance accessibility and fairness while maintaining the validity of test scores.

In national assessment systems, accommodations have become a central component of inclusive education policies. Common practices include extended testing time, additional breaks, alternative formats (e.g., Braille or large print), assistive technologies, and adapted testing environments (Educational Testing Service, 2024). Research suggests that such accommodations can enable students with diverse needs to participate meaningfully in assessments without compromising measurement integrity (Thompson et al., 2002; Thurlow et al., 2001).

In addition, the principles of Universal Design (UD) have increasingly been discussed in relation to assessments as a way of promoting accessibility from the outset rather than addressing it only through accommodations introduced after assessment design (Christensen et al., 2014; Dolan et al., 2010; Thurlow, 2021; World Bank, 2023).

In the field of international assessments, the OECD has developed guidelines for revising and designing PISA items aimed at enhancing accessibility, based, among other frameworks, on UD principles (Voisin et al., 2025). The IEA has also incorporated accessibility considerations into its updated Technical Standards, which now include a standard on “Ensuring Accessibility Support” (Martin et al., 2025). This standard provides guidelines for the planning and design of studies, including references to UD principles in instrument development to promote accessibility for the broadest possible range of participants.

Although efforts to incorporate UD principles are increasing, accommodations are currently the main approach to supporting participation in international assessments. The IEA Technical Standards also recognize accommodations as one component of accessibility support, and IEA study manuals, such as the Survey Operations Procedures and the School Coordinator Manual of current

and recent study cycles provide some guidance on the types of accommodations that may be permitted during testing. Table 1 provides a summary of the accommodations mentioned in these manuals across the latest completed cycles of IEA studies.

Table 1

Overview of Accommodation Guidance Across IEA Study Manuals

Study	General guidance on accommodations	Accommodations mentioned in study manuals
ICCS 2022	Not identified in available documentation	Not identified in available documentation
ICILS 2023	Accommodations should be consistent with national policies and procedures for supporting students with special educational needs.	Additional time (disabling timer); separate testing room; human reader; human scribe; screen magnification software; additional breaks; administration across two days.
PIRLS 2021	Accommodations should be provided in accordance with national and school policies for students who are conceptually capable of participating in the assessment but require support to access it (e.g., due to hearing, visual, dyslexia, or physical impairments).	Disabling automatic timer; separate testing room with an additional test administrator; printed materials on a larger paper in larger print; use of a magnifying glass. Reading passages/items aloud is not permitted.
TIMSS 2023	Accommodations should be provided in accordance with national and school policies for students who are conceptually capable of participating in the assessment but require support to access it (e.g., due to hearing, visual, dyslexia, or physical impairments).	Disabling automatic timer; separate testing room with an additional test administrator; printed materials on a larger paper in larger print; use of a magnifying glass.

Note. Sources refer to the Survey Operations Procedures and School Coordinator Manual of each study cycle (ICCS 2022: IEA, 2021a, 2021b; ICILS 2023: IEA, 2022a, 2022b; PIRLS 2021: IEA, 2019, 2020; TIMSS 2023: IEA, 2021c, 2022c).

Although study manuals provide some guidance on the types of accommodations that may be permitted during testing, systematic information on how accommodations are implemented across countries remains limited. Technical documentation of IEA studies focuses primarily on sampling and implementation procedures (LaRoche et al., 2020; Almaskut et al., 2023), providing limited insight

into how accommodations are applied in practice. Recent studies have started to address this gap. For example, as part of the international quality control program of IEA assessments, some information is collected on whether special accommodations were provided during samples of testing sessions in each country, with a few more details on specific accommodations used reported for ICCS 2022 and ICILS 2023 (see Musu & Hartmanová, 2024; Hartmanová et al., 2025). However, more detailed and systematic comparable evidence is missing across all IEA assessments. In contrast, emerging work in the PISA context has begun to address this gap by mapping accommodation practices more systematically and linking them to specific student needs (Guez et al., 2024).

Study Aim and Contribution

This study examines how students with different difficulties are included or excluded and accommodated in international large-scale assessments conducted by the IEA. Moving beyond prior research that has primarily focused on documenting exclusion rates, the study adopts an implementation-oriented perspective, focusing on the processes and decision-making practices that shape participation in international assessments. In doing so, it seeks to uncover not only the extent of inclusion or exclusion, but also the underlying mechanisms through which these outcomes are produced across different national contexts.

Drawing on cross-national survey data from national research coordinators responsible for implementing TIMSS, PIRLS, ICCS, and ICILS, the study provides a previously undocumented perspective on how participation decisions are made at the national level. This approach enables a systematic examination of how eligibility criteria are interpreted in practice, how various difficulty categories influence inclusion or exclusion decisions, and how accommodations are applied across diverse educational systems. By focusing on implementation processes rather than solely outcomes, the study contributes to a more nuanced and policy-relevant understanding of participation in international assessments.

Specifically, the study addresses the following research questions:

- (1) In which educational settings are students with different types of difficulties schooled, and how does this relate to their representation in international assessments?

- (2) Which types of difficulties are most frequently associated with student inclusion or exclusion from IEA assessments, and who is involved in making these decisions?
- (3) What accommodations are provided to students who participate, and what resources do countries identify as necessary for greater inclusion?

Methods

This study employs an international comparative approach to examine exclusion and accommodation practices applied by national study centers in educational assessments conducted by the IEA. Systematic cross-national comparison establishes evidence on the current state of practice and reveals the range of international variation in the implementation of exclusions and accommodations across countries participating in IEA studies. Documenting this variation is valuable because it makes visible what is possible: countries differ substantially in the extent to which they include and accommodate students, and these differences can inform efforts to expand participation. The survey instrument combines closed and open-ended items, enabling a mixed-methods analytic strategy that pairs descriptive quantitative analysis with qualitative content analysis of open-ended responses. Full methodological details, including the classification frameworks, questionnaire structure, sample composition, and analytic procedures, are provided in Supplementary Material A.

Participants and Data Collection

We surveyed national research coordinators responsible for the national implementation of five IEA assessments: TIMSS 2023 (Grades 4 and 8), PIRLS 2026 (grade 4), ICCS 2022 (grade 8), and ICILS 2023 (grade 8). National research coordinators oversee sampling, translate exclusion criteria to the national context, and coordinate accommodations, making them the most knowledgeable informants on how international assessment guidelines are translated into national practice. For PIRLS, national research coordinators were surveyed during the preparation phase of the 2026 cycle and reported on practices that took place during the field trial.

In total, 234 country-study survey invitations were distributed to national research coordinators across 85 educational systems, including both countries and subnational entities (e.g.,

Ontario, Dubai, England). Of these, 135 responses were received (response rate: 58%), covering 63 countries and subnational entities. The sample spans Europe, the Middle East and North Africa, the Asia-Pacific region, the Americas, and Sub-Saharan Africa, though European systems are overrepresented (see Supplementary Material A for the full country list and assessment-level sample sizes). National research coordinators were contacted by email and invited to participate in the online survey. Two reminders were sent to non-respondents during the data collection period. The survey was administered between 5th of July and 15th of October 2025.

Instrument

The questionnaire is organized around two classification frameworks. The first is a typology of 17 functional difficulties that may affect assessment participation, spanning intellectual disabilities, physical and sensory impairments, specific learning difficulties, emotional and behavioral challenges, chronic health conditions, and limited proficiency in the language of assessment. As outlined in the introduction, we focus on difficulties rather than by diagnostic categories (see Supplementary Material A for details). The second is an inventory of 39 accommodations organized into five domains: presentation (14), response (10), setting (7), timing and scheduling (5), and medical assistance (3). For each of the 17 difficulty types, national research coordinators indicated which of the 39 accommodations their country provides, yielding a comprehensive cross-tabulation of national practices. Beyond accommodations, national research coordinators reported on educational placement, inclusion or exclusion rates, decision-making processes, and the resources needed for greater inclusion (see Supplementary Material A for full details on instrument development and item content).

Analysis

For closed-ended items we computed frequency counts and cross-tabulations; for open-ended responses we applied qualitative content analysis. All analyses were conducted at the country–study level. Because many countries contribute multiple entries in the merged dataset (with survey responses for all IEA studies), we applied inverse frequency weights so that each country contributes

equally to aggregate statistics regardless of the number of survey entries it contributes (e.g., a country appearing twice in the dataset received a weight of 0.5 per response). The analytical sample consists of 63 countries, which form the basis for all percentage distributions reported in the figures, unless otherwise specified.

Results

The results are organized around four key aspects of inclusion for international assessments. First, we examined where students with difficulties are most commonly educated, either in special or in mainstream schools. Second, we investigated the extent to which special schools are included in or excluded from international assessments. Next, we turn to student-level exclusions across the 17 difficulty types identified in this study, examining for each difficulty type the proportion of students who are included and how these decisions are made. Finally, we report the accommodations provided to students included in the assessments and examine how frequently each accommodation type is used across countries, by difficulty type.

School System Inclusiveness: Where Do Students with Difficulties Go to School?

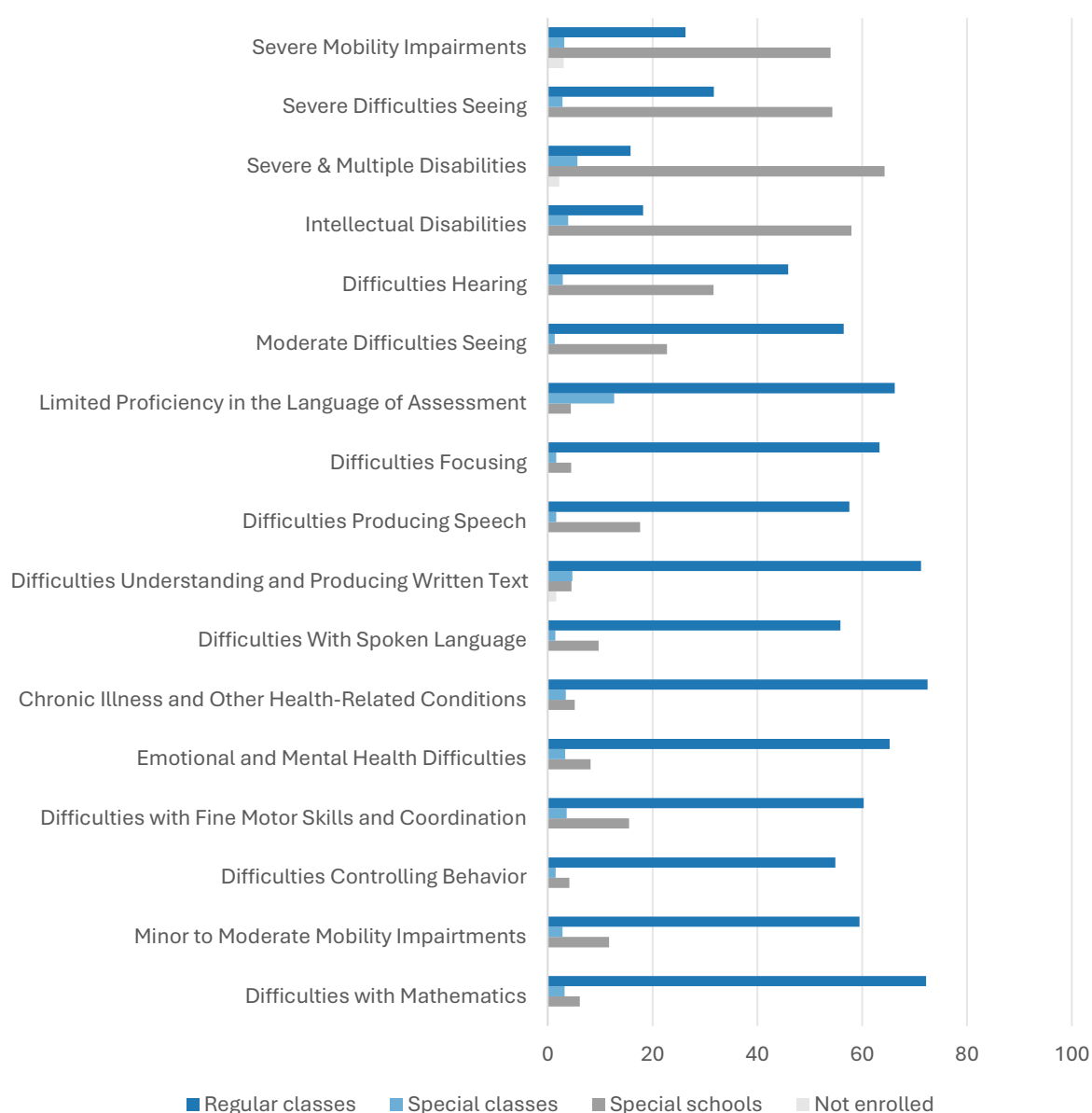
School placements for students with difficulties may vary across education systems and according to the type of difficulty. These placements can include education in regular classes within mainstream schools (often referred to as inclusive schooling), special classes within mainstream schools (referred to as separated schooling), and placement in special schools (referred to as segregated schooling) (Snozzi et al., 2025). These arrangements reflect different approaches to organizing educational provision for students with diverse needs, with some approaches advocating education for all students in mainstream classrooms and others emphasizing the availability of a range of placement options (Hornby, 2014).

We asked how students with different difficulties are schooled in each country. For the sake of completeness, we also asked whether students with certain difficulties were not enrolled in school at all. Distinct patterns of school placement emerge across difficulty categories. Figure 1 presents the weighted percentages of responses indicating where “all or almost all” students attend school, by difficulty type. Students with more severe or complex needs, such as those with severe and multiple

disabilities or intellectual disabilities, are more frequently placed in special schools, while placement in regular classes in mainstream school is comparatively lower for these groups. However, the presence of placement in regular classes shows that inclusive practices exist to some extent in some of the educational systems in our sample. For most other difficulties, placement in regular classes in mainstream schools remains the most common arrangement, particularly for students with less severe or more specific difficulties (e.g., learning, language, or behavioral difficulties).

Figure 1

Percentage of countries reporting where “all or almost all” of students go to school, by difficulty type.



Note. Percentages are based on the 63 countries included in the analytical sample.

Exclusions of Special Schools

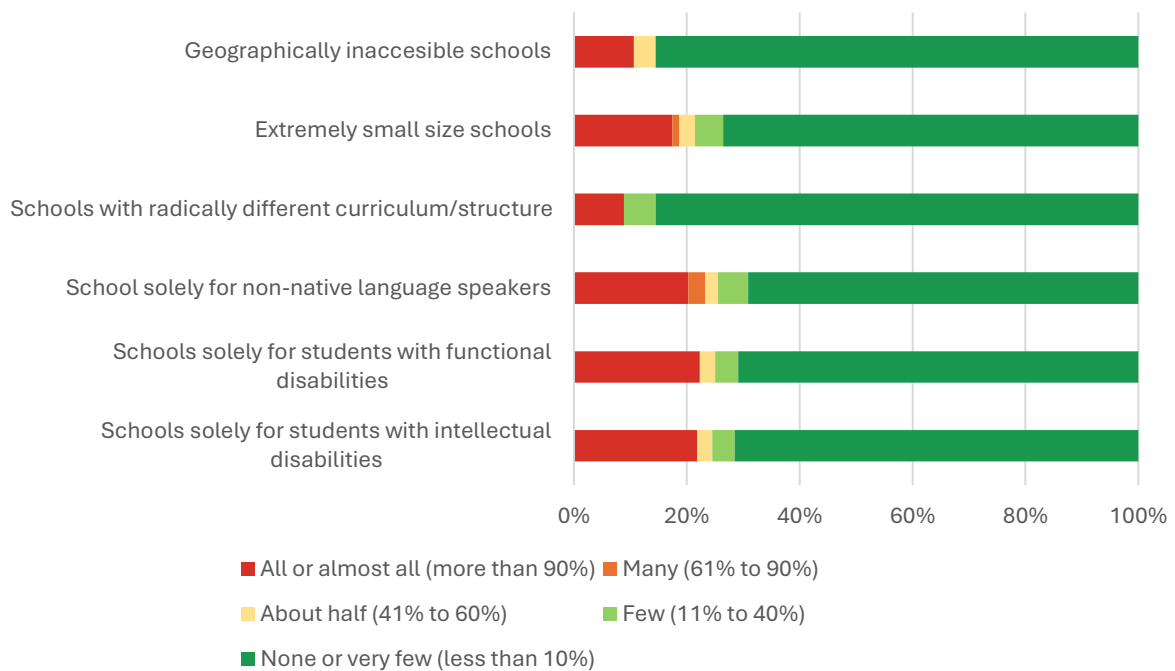
Figure 2 shows the extent to which different types of schools are excluded in international assessments, focusing only on the types of exclusions explicitly permitted in the four studies examined in this study. These exclusion categories do not relate exclusively to special schools; however, all are included here for comparison purposes. The six categories can be grouped into structural or logistical (e.g., geographically inaccessible or very small schools) and special school exclusions (schools serving only students with intellectual or functional disabilities or only non-native language speakers). This distinction allows comparison of whether school exclusions are driven primarily by practical constraints or by the nature of the schools themselves.

Notably, the special school category in these studies captures only schools serving students with functional or intellectual disabilities or non-native language speakers and does not encompass all the difficulties considered in this study. Nevertheless, the results presented in Figure 2 show that, in general, schools serving only students with special needs are more frequently excluded in international assessments, with the largest reports of excluding “about half”, “many” or “all or almost all” of these schools. Given that students with mainly the severe or complex difficulties, tend to attend these schools (see previous section), these students are particularly likely to be excluded at the school selection stage.

In contrast, schools are less frequently excluded for more practical or logistical reasons, such as geographic location or having a radically different curriculum. Only schools of extremely small size are excluded to a similar extent as special schools. Overall, the results show exclusion practices remain limited but are somewhat more frequent for schools defined by student characteristics than for those defined by structural or geographic factors.

Figure 2

Extent of school-level exclusions in ILSAs, by type of school

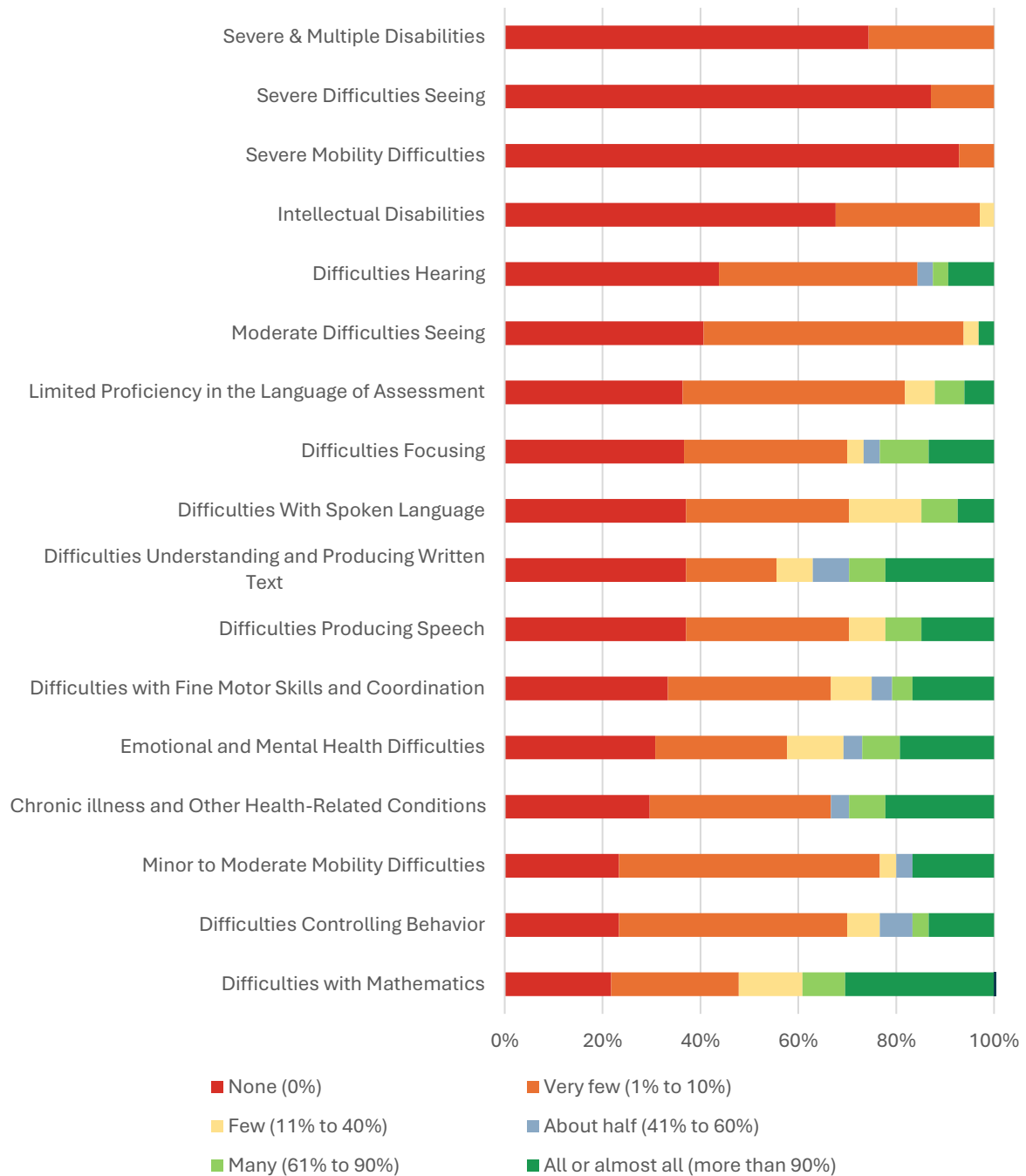


Student-Level Inclusion by Difficulty Type

This section reports the proportion of affected students who were excluded from the assessments for each of the 17 difficulty categories. Figure 3 presents the percentages of countries reporting different levels of student inclusion in international assessments for each difficulty type. The results show some variation in inclusion practices, but still a consistent pattern emerges among the countries that provided data: very few students, if not none, are included in international assessments. However, the degree of inclusion is also linked to the type and intensity of support needs. For students with more complex or intensive needs, such as severe and multiple disabilities, severe difficulties seeing, severe mobility difficulties, and intellectual disabilities, countries overwhelmingly report that “none” or “very few” students are included in the assessments. In contrast, higher and more variable levels of inclusion are observed for students with more specific learning or behavioral difficulties, such as difficulties with mathematics, difficulties understanding or producing written text, or emotional or mental health difficulties, where a larger share of countries report including “many” or “all/almost all” students.

Figure 3

Share of countries reporting how many students are included in international assessments, by difficulty type



Note. The figure shows distributions based on valid responses only. “Data not available” and item non-response are excluded, and percentages are rescaled to 100% within each difficulty. Depending on the difficulty, approximately 30–50% of country-level observations are excluded due to missing item-level information.

Another important observation concerns the variation observed within each difficulty category. Any response other than “none” indicates that some students are included while others are excluded, albeit in different proportions. For example, although “very few,” “about half,” and “many” each describe different levels of inclusion, they all point to selective inclusion: decisions are being made about which students with a given type of need are allowed to participate and which are not.

A Qualitative Analysis of Decision-Making Processes for Student Inclusion and Exclusion

To better understand this process, national research coordinators were asked to describe how these decisions are made, including who was involved and what type of input or documentation is required. Their responses were analyzed independently by two raters using qualitative content analysis. In case of a disagreement, the raters were consulted to reach a consensus on the coding. The results show that across all types of difficulties, decisions on students’ inclusion or exclusion were predominantly made at the school level by teachers (range: 33–42 mentions, depending on the difficulty), school coordinators (26–31 mentions), or school principals (19–27 mentions). In many cases, NRCs did not specify the required input or documentation (category count range: 30–45, depending on the difficulty). Where information was provided, the most frequently cited requirement was an official medical diagnosis (19–39 mentions across difficulties). This requirement was reported less often for “Limited proficiency in the language of assessment” (19 mentions) and “Emotional and mental health difficulties” (27 mentions). Another common response category indicated that in some countries and studies no formal documentation was required (13–16 mentions, depending on the difficulty).

Mapping of Accommodations by Difficulty Type

Table 2 shows the proportion of countries reporting the use of each assessment accommodation for each difficulty category. Darker blue shading indicates a higher share of countries reporting that the accommodation is provided to students. The number of countries included varies by difficulty category, as countries were only asked about assessment accommodations if they reported at

least some level of inclusion for the respective difficulty category. The first row indicates the number of countries contributing to the proportion for each difficulty category.

The results show clear variation in the use of accommodations. Timing and scheduling accommodations, setting accommodations, and medical accommodations are among the most frequently reported across difficulty categories. For example, extended time is reported by 25-45% of countries across difficulties categories, with the highest reported use for students with moderate difficulties seeing (45%) and difficulties focusing (43%). Individual testing is reported by 21-36% of countries, with the highest reported use for students with difficulties controlling behavior (36%). In contrast, more specialized accommodations, particularly those within the response and presentation categories, are reported far less frequently or not at all. For example, the use Braille test materials is rarely reported, even among the categories involving moderate and severe difficulties seeing, the groups for which this accommodation is most applicable. Similarly, adapted texts and text-to-speech software are rarely reported (below 7% across categories). An exception is the provision of a human reader or a human scribe, which are both among the most frequently used accommodations (weighted averages 24% and 21%, respectively).

The pattern of accommodations provided also varies systematically by difficulty type. Students with moderate difficulties seeing or with difficulties focusing are more likely to have access to multiple accommodations, while students with more complex or severe difficulties, such as severe difficulties seeing or severe mobility difficulties, are offered fewer supports. A similar pattern is observed for students with limited proficiency in the language of assessment, who are also reported to receive fewer supports. The results further show that, among countries reporting some level of inclusion for each difficulty category, a substantial proportion report providing no accommodations (31–49%, depending on the difficulty category).

All in all, these findings show that although some students with different difficulty categories are included in international assessments, their participation may remain limited due to limited provision of necessary accommodations.

Table 2
Mapping of accommodations by difficulty

		Intellectual Disabilities	Moderate Difficulties Seeing	Severe Difficulties Seeing	Difficulties Hearing	Minor Mobility Impairments	Severe Mobility Impairments	Fine Motor/Coordination Difficulties	Difficulties Producing Speech	Difficulties Understanding Written Text	Difficulties with Mathematics	Difficulties Controlling Behavior	Difficulties Focusing	Emotional Difficulties	Chronic Illness/Health Conditions	Limited Language Proficiency	Severe/Multiple Disabilities	Difficulties With Spoken Language	WEIGHTED AVERAGE
	N (weighted)	25	37	21	30	40	20	36	33	34	38	37	32	36	36	34	28	34	
	No accommodations provided	.42	.31	.42	.45	.37	.41	.41	.49	.38	.44	.35	.36	.49	.38	.45	.34	.32	.40
Presentation	Adjustments to print materials	.04	.14	.05	.07	.09	.00	.11	.11	.04	.12	.08	.08	.07	.07	.04	.07	.07	.08
	Adjustments to digital interface	.11	.20	.13	.06	.08	.07	.12	.10	.07	.12	.09	.11	.09	.09	.06	.12	.10	.10
	Magnification devices	.08	.18	.10	.04	.05	.08	.06	.04	.03	.06	.03	.04	.05	.03	.00	.08	.03	.06
	Adapted texts	.00	.05	.00	.03	.03	.00	.03	.00	.00	.05	.00	.00	.00	.00	.00	.07	.01	.02
	Simplified supports	.00	.05	.00	.03	.06	.00	.03	.00	.00	.03	.00	.00	.00	.00	.00	.04	.00	.02
	Braille test materials	.00	.00	.00	.00	.00	.00	.03	.00	.00	.03	.00	.00	.00	.00	.00	.00	.00	.00
	Screen reader software/Text-to-speech software	.02	.02	.03	.01	.01	.02	.04	.01	.02	.03	.00	.01	.01	.01	.01	.06	.02	.02
	Human reader	.35	.32	.35	.18	.23	.18	.24	.18	.35	.25	.20	.24	.20	.16	.16	.26	.20	.24
	Human translator	.06	.01	.00	.02	.04	.00	.04	.03	.00	.03	.00	.00	.00	.01	.17	.04	.01	.03
	Text translation (printed or digital text translation into another language)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.03	.00	.00	.00	.00	.03	.00	.03	.01
	Recorded instructions	.00	.00	.00	.00	.00	.00	.03	.00	.00	.03	.00	.00	.00	.00	.03	.00	.00	.01
	Sign language interpretation (live or recorded)	.00	.00	.00	.11	.00	.00	.03	.06	.00	.03	.00	.00	.00	.00	.03	.00	.06	.02
	Cued speech (use of hand cues to show speech sounds)	.00	.00	.00	.03	.00	.00	.03	.03	.00	.03	.00	.00	.00	.00	.03	.00	.03	.01
	Audio amplification devices (e.g., FM systems)	.00	.00	.00	.06	.00	.00	.03	.00	.00	.03	.00	.00	.00	.00	.00	.02	.00	.01
Response	Human writer/scribe	.23	.31	.23	.17	.23	.26	.26	.20	.26	.16	.16	.16	.16	.17	.20	.27	.23	.21

	Speech-to-text technology (voice recognition software)	.02	.03	.02	.00	.00	.00	.03	.00	.01	.03	.00	.00	.00	.00	.00	.01	.00	.01
	Word processor (typing responses on a computer)	.09	.05	.09	.06	.08	.08	.08	.02	.06	.07	.02	.05	.02	.05	.02	.02	.09	.05
	Braille writer/keyboard	.00	.01	.00	.00	.00	.00	.03	.00	.00	.03	.00	.00	.00	.00	.00	.00	.00	.00
	Calculators (basic, specialized, or four-function calculator use)	.02	.04	.03	.02	.02	.03	.04	.02	.02	.07	.02	.02	.02	.02	.01	.02	.02	.03
	Spelling and grammar check (digital tools)	.00	.03	.00	.03	.00	.00	.00	.03	.00	.03	.00	.00	.00	.00	.01	.04	.03	.01
	Highlighters (digital or paper-based)	.04	.05	.05	.00	.03	.05	.03	.00	.03	.03	.00	.03	.00	.00	.00	.02	.03	.02
	Specific keyboard or mouse (adaptive input devices)	.16	.09	.19	.07	.12	.18	.16	.11	.13	.11	.09	.11	.07	.09	.06	.12	.10	.11
	Specific writing materials (e.g., pencil grips, slant boards, special pens)	.00	.01	.07	.00	.01	.00	.07	.03	.03	.05	.03	.03	.01	.03	.03	.01	.03	.03
	Assistive technology (e.g., eye-control, breath-control)	.00	.00	.00	.03	.03	.02	.03	.04	.00	.03	.03	.00	.03	.00	.00	.00	.01	.01
Setting	Small group testing	.28	.27	.17	.22	.20	.17	.19	.29	.24	.21	.24	.26	.23	.21	.18	.26	.22	.23
	Individual (one-on-one) testing	.30	.35	.23	.29	.28	.21	.32	.30	.26	.25	.36	.32	.28	.23	.26	.29	.29	.29
	Preferential seating	.28	.24	.22	.23	.29	.23	.32	.24	.25	.21	.31	.33	.24	.27	.17	.16	.27	.25
	Wheelchair-accessible testing locations	.16	.18	.18	.11	.40	.24	.17	.14	.10	.13	.12	.14	.11	.14	.07	.09	.15	.16
	Noise buffers (e.g., earplugs, noise-canceling headphones)	.15	.10	.11	.08	.11	.13	.10	.10	.10	.07	.09	.12	.09	.09	.07	.10	.12	.10
	Changes to room conditions (e.g., lighting, temperature, furniture)	.18	.18	.13	.12	.19	.13	.15	.15	.11	.10	.17	.16	.18	.18	.10	.10	.13	.15
	Minimal visual/auditory distractions	.13	.14	.21	.11	.12	.16	.18	.14	.10	.11	.18	.20	.14	.15	.10	.08	.17	.14
Timing and scheduling	Multiple-day testing	.06	.08	.03	.07	.04	.05	.06	.09	.03	.02	.07	.08	.04	.04	.09	.09	.03	.06
	Flexible scheduling	.09	.16	.09	.13	.14	.10	.17	.13	.11	.14	.15	.17	.10	.14	.15	.17	.14	.14
	Extended time	.42	.45	.25	.35	.36	.26	.33	.29	.41	.36	.38	.43	.32	.35	.30	.36	.30	.35
	Additional or extended breaks	.29	.26	.30	.26	.27	.24	.24	.22	.18	.14	.31	.26	.26	.26	.10	.28	.17	.23
	Shorter test	.00	.03	.00	.00	.01	.02	.00	.00	.01	.00	.00	.01	.00	.01	.00	.04	.00	.01
Medical	Permission for food/drink during testing	.32	.32	.32	.27	.25	.29	.31	.27	.23	.24	.29	.32	.28	.36	.15	.30	.30	.28
	Permission for medication or medical devices use during testing	.28	.28	.25	.25	.22	.29	.27	.26	.19	.21	.26	.28	.24	.37	.15	.20	.28	.25
	Monitoring by medical personnel	.12	.11	.15	.12	.11	.18	.10	.14	.07	.08	.14	.12	.12	.14	.08	.14	.10	.12
Other (not listed)		.13	.06	.17	.10	.11	.13	.07	.11	.08	.06	.13	.11	.11	.07	.06	.13	.10	.10

Note. N (weighted) indicates the number of countries included in each difficulty category. Values in the table represent the proportion of these countries reporting the provision of each accommodation.

Perceived Resource Needs for Enhancing Accessibility in Educational Studies

To better understand which resources could enhance accessibility in educational assessments, our survey included an open-ended question asking “Which resources or forms of support could help IEA make its studies more accessible?”. Responses to this question were analyzed using qualitative content analysis. The most frequently mentioned category was “*Collaboration and Feedback*” (36 mentions), including best practice sharing and expert groups. The second category was “*Guidelines and Policy*” (26 mentions), referring to formal or semi-formal guidance, inclusion and exclusion criteria, ethical considerations, and regulations on permitted accommodations. Another frequently mentioned response was the need for workshops or formal training on inclusive assessment practices in IEA studies (19 mentions). Notably, all three categories reflect a demand for additional information and structured support rather than for technical tools or financial resources. This suggests that inclusive assessments are not yet fully understood or consistently applied across countries.

Discussion

This study examined how countries define, include or exclude, and accommodate students with difficulties in international assessments. Drawing on country-weighted evidence across multiple IEA studies, the findings show that participation is not evenly distributed across difficulty categories but is structured by the type of difficulty, the educational settings in which students are located, and the kinds of accommodations that countries are prepared to provide. Taken together, the results suggest that who is represented in international assessments is shaped not only by formal eligibility rules but also by broader institutional arrangements and practical implementation decisions.

A first key finding is that exclusion is strongly patterned by the type of difficulty. Students with severe and multiple disabilities, severe mobility impairments, severe difficulties seeing and intellectual disabilities were among the most excluded groups, whereas students with language, learning, and behavioral difficulties were comparatively more likely to be included. This extends previous research showing cross-national variation in exclusion rates in international assessments (Atasever et al., 2024; Kjeldsen & Strietholt, 2024). More importantly, the findings suggest that exclusion decisions are not determined by difficulty type alone. Rather, they reflect how participation

criteria are interpreted and applied in practice, likely shaped by a combination of formal guidance, diagnostic classifications, and contextual factors such as resources and available accommodations. As a result, inclusion varies not only across difficulty categories but also within them, even within the same country. In this sense, the results support concerns that international comparisons may partly reflect differences in participation practices rather than only differences in student achievement (Klieme, 2012; Meinck, 2020). From a validity perspective, this matters because systematic underrepresentation of specific groups can distort the interpretation of national performance estimates and weaken claims that assessment results fully represent target populations (Messick, 1989; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

A second important finding is that educational placement appears to be closely linked to assessment participation. Students with more severe or complex needs were more often reported to be located in special schools or segregated settings, while students with language, learning, and behavioral difficulties were more frequently situated in regular classes in mainstream schools. This distinction highlights that exclusion operates both at the level of assessment decisions and at the level of educational system organization. Where students with SEN are educated in settings outside the assessment's sampling frame, they may be effectively excluded before any participation decision is made. This finding is consistent with prior work showing that national differences in SEN classification and educational provision shape the visibility of these students in assessment systems (Schuelka, 2013; Florian, 2008). It also reinforces the point that participation patterns in international assessments cannot be fully understood without considering institutional arrangements in special and inclusive education.

A third contribution of the study concerns assessment accommodations as a mechanism for inclusion. The findings show that accommodations are used across countries, but not equally across types of support. Timing-related accommodations, such as extended time, additional breaks, and flexible scheduling, as well as setting-related adjustments such as individual or small-group testing, were the most frequently reported. By contrast, more specialized presentation and response accommodations, including Braille materials, screen readers, speech-to-text tools, and other assistive

technologies, were much less common. This pattern is in line with the broader accommodations literature, which shows that schools and testing systems tend to rely more heavily on flexible administrative adjustments than on highly specialized supports that require technical infrastructure or individualized adaptation (Fuchs & Fuchs, 1999; Lovett & Lewandowski, 2015). At the same time, this uneven distribution is consequential. The students most likely to require specialized accommodations are also those who appear least likely to be included. Thus, accommodations appear to facilitate participation for some groups, but they do not fully offset barriers for students with more complex sensory, physical, or multiple disabilities.

These findings highlight a broader tension between the goals of comparability, feasibility, and inclusion in international assessments. IEA studies, like other international assessment programs, seek to standardize administration conditions to preserve comparability across countries (Wagemaker, 2020). Yet the present results suggest that this standardization may come at the cost of excluding students whose participation would require more substantial adaptation. This tension is not unique to IEA; it is a longstanding issue in educational measurement, where efforts to maintain score comparability must be balanced against the need to reduce construct-irrelevant barriers (AERA, APA & NCME, 2014; Messick, 1989). The findings therefore suggest that inclusion in international assessments should not be treated only as an operational issue but as a core component of assessment validity and fairness.

The study also speaks directly to the problem of definitional ambiguity in difficulty categories. As the literature has noted, special need, difficulty or disability are not fixed or universally applied constructs; rather, it varies across legal systems, diagnostic traditions, and educational policies (Schuelka, 2013; Mezzanotte, 2023). The cross-national variation observed here in inclusion, exclusion, and accommodation patterns is consistent with the argument that broad categories such as intellectual disabilities, functional disabilities, or language barriers leave substantial room for interpretation during implementation. In practice, this means that countries may be working with formally similar categories but applying different thresholds and expectations when deciding who can participate. Without clearer operational guidance, international assessments risk comparing not only achievement outcomes but also national classification systems and local judgments about

participation. The framework adopted in this study aims to address some of these challenges by providing a broader and more flexible approach to categorizing difficulties, which may support more consistent interpretation across education systems.

More broadly, the findings raise important questions for the educational research community about the representativeness of the evidence base that informs policy. If the assessments used to monitor progress toward SDG 4—a goal that explicitly aims to ensure inclusive and equitable quality education for the most marginalized learners (UNESCO, 2015)—systematically exclude students with difficulties, a fundamental circularity problem arises: the populations that inclusive education policies are designed to serve are absent from the data used to evaluate whether those policies are working. This concern extends beyond international assessments to national assessments, program evaluations, and school-based research studies where sampling frames, standardized instruments, and testing conditions may replicate similar exclusion patterns without systematic documentation or scrutiny (Thurlow et al., 2001; Ketterlin-Geller, 2008).

The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) require that test developers consider the full population of intended test-takers and design for accessible participation as a matter of validity. Yet these principles are applied unevenly—more rigorously in high-stakes national assessment programs, where legal mandates such as the Individuals with Disabilities Education Improvement Act (2004) require participation, than in the international comparisons and educational research studies that shape global policy discourse. The movement toward UD for Assessment offers a constructive framework for addressing this gap (Thompson et al., 2002; Thurlow, 2021), but its application in international assessment contexts remains limited, although it is increasingly reflected in more recent guidance from international assessment bodies, including the updated IEA Technical Standards on Ensuring Accessibility Support (Martin et al., 2025). The present findings suggest that international assessment organizations and the broader educational research community would benefit from adopting inclusive assessment design as a core methodological standard, moving beyond post-hoc accommodation to proactive accessibility.

From a policy and methodological perspective, the findings underscore the need for greater transparency in documenting participation practices. Current international documentation often

indicates whether accommodations were used, but provides limited information about which accommodations were offered, to whom, and under what conditions (Martin & Mullis, 2012; LaRoche et al., 2020). More systematic reporting would improve the interpretability of international results and help users assess how representative national samples actually are. In addition, the results suggest that future development work in IEA assessments should invest more seriously in accessible assessment design, including the implementation of UD principles, as outlined in the IEA Technical Standards, as well as the validation of specialized accommodations. This concerns, on the one hand, the use of targeted accommodations that match the specific difficulties and, on the other hand, the evaluation of whether these accommodations actually prove to be effective. Another aspect that could be further strengthened in IEA studies is the provision of more explicit guidance for national research coordinators regarding inclusion and the use of accommodations, including both clear information on who is eligible to receive accommodations and which accommodations are permitted or not permitted within each study. Recent efforts in other international contexts to map accommodations more systematically point in this direction (Guez et al., 2024).

Limitations

Some limitations should be noted. First, the study relies on survey responses from national research coordinators, which means the results depend on the accuracy and consistency of respondent reports. Second, we opted to summarize results at the country level rather than analyze individual studies separately. This required a country-weighting approach to avoid overrepresenting countries participating in multiple studies. While this approach supported a broader cross-national overview, it limits the extent to which study-specific patterns can be examined. Future research could examine the consistency of reported practices across studies, as well as differences across subject domains or target grades. A third limitation concerns the overall response rate and the availability of data for some key variables. Potential bias may have been introduced if countries that responded, or that were able to provide the requested information, differ systematically from those that did not. However, no follow-up data were collected from non-participating countries to examine the reasons for non-response, and the findings should therefore be interpreted cautiously. In many cases, missingness

reflected the unavailability of detailed information at the national level rather than simple item non-response. Nevertheless, the study provides valuable comparative evidence on inclusion and accommodation practices across a wide range of education systems, while also highlighting the challenges of collecting internationally comparable information in this area. Frameworks such as the one used in this study, which conceptualize difficulties in terms of functional needs and participation rather than fixed diagnostic categories, may help shift the focus from diagnostic identification toward information about students' participation and support needs.

Conclusion

Despite these limitations, the study makes an important contribution to understanding assessment participation as a substantive issue of educational equity and measurement validity—not merely a technical detail of survey administration. The findings demonstrate that some groups of students, particularly those with severe, physical, or sensory disabilities, remain systematically less visible in international assessments, while other groups are more likely to be included and supported through accommodations. Importantly, these findings suggest that inclusion should not be understood merely as formal participation, but as meaningful participation supported by appropriate accommodations. These patterns have implications that extend well beyond IEA assessments. They speak to longstanding concerns in the measurement literature about the validity of population-level inferences when identifiable subgroups are systematically excluded (Messick, 1989; AERA, APA, & NCME, 2014), to the coherence of global education monitoring frameworks that rely on assessment data to track progress toward inclusive education goals (UNESCO, 2015, 2020), and to the responsibility of the educational research community to ensure that the populations it studies reflect the diversity of the populations it seeks to serve. As the Salamanca Statement declared three decades ago, “inclusion and participation are essential to human dignity and to the enjoyment and exercise of human rights” (UNESCO, 1994, p. 11). Extending this principle from education to the assessments and research that inform education remains an urgent and unfinished task.

References

- Almaskut, A., LaRoche, S., & Foy, P. (2023). Sample design in PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and procedures: PIRLS 2021 technical report* (pp. 3.1–3.32). TIMSS & PIRLS International Study Center, Boston College.
<https://doi.org/10.6017/lse.tpisc.tr2103.kb9560>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andersson, C., & Sandgren Massih, S. (2023). PISA 2018: Did Sweden exclude students according to the rules? *Assessment in Education: Principles, Policy & Practice*, 30(1), 33–52.
- Atasever, U., Jerrim, J., & Tieck, S. (2024). Exclusion rates from international large-scale assessments: An analysis of 20 years of IEA data. *Educational Assessment, Evaluation and Accountability*, 36(3), 405–428.
- Aursand, L., & Rutkowski, D. (2021). Exemption or Exclusion? A study of student exclusion in PISA in Norway. *Nordic Journal of Studies in Educational Policy*, 7(1), 16-29.
- Bjerril, S. (2024, March 15). Hvor misvisende var det danske PISA resultat? Nyt studie tyder på, at det kan være rigtig meget. *Folkeskolen*. <https://www.folkeskolen.dk/finland-matematik-norge/hvor-misvisende-var-det-danske-pisa-resultat-nyt-studie-tyder-pa-at-det-kan-vaere-rigtig-meget/4760039>
- Brussino, O. (2020). Mapping policy approaches and practices for the inclusion of students with special education needs (OECD Education Working Papers, No. 227). OECD Publishing.
<https://doi.org/10.1787/600fbad5-en>
- Christensen, L. L., Shyyan, V., & Johnstone, C. (2014). Universal design considerations for technology-based, large-scale, next-generation assessments. *Perspectives on Language and Literacy*, 40(1), 23–31.
- College Board. (2025). *Spring 2025 accommodations and supports handbook*. College Board.
- Committee on the Rights of Persons with Disabilities. (2016). *General comment No. 4 (2016), Article 24: Right to inclusive education* (CRPD/C/GC/4). United Nations.

<https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-4-article-24-right-inclusive>

- Cortes, D., & Atasever, U. (2024). Sampling design and implementation. In W. Schulz, T. Friedman, & J. Fraillon (Eds.), *ICCS 2022 technical report* (pp. 73–82). International Association for the Evaluation of Educational Achievement (IEA).
- Dolan, R.P., Burling, K.S., Rose, D., Beck, R., Murray, E., Strangman, N., Jude, J., Harms, M., Way, W.D., Hanna, E. and Nichols, A., (2010). Universal design for computer-based testing (UD-CBT) guidelines. Pearson. <https://doi.org/10.13140/RG.2.2.31923.35361>
- Educational Testing Service. (2024). *Disabilities and health-related needs: Common testing accommodations*. <https://www.ets.org/disabilities/test-takers/common-testing-accommodations.html>
- Florian, L. (2008). Inclusion: Special or inclusive education: Future trends. *British Journal of Special Education*, 35(4), 202–208.
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School Administrator*, 56(10), 24–29.
- Guez, A., Ketan, A., & Piacentini, M. (2024). *Mapping study for the integration of accommodations for students with special education needs (SEN) in PISA* (OECD Education Working Paper No. 308). OECD Publishing. <https://doi.org/10.1787/ed03c717-en>
- Hartmanová, K., Musu, L., & Moreno Hidalgo, M. (2025). Quality assurance in the ICILS data collection. In J. Fraillon, M. Rožman, S. Meyer, L. Musu, Y. Liaw, A. Christiansen, & S. Tieck (Eds.), *ICILS 2023 Technical Report* (pp. 125 – 156).
- Hornby, G. (2014). *Inclusive special education: Evidence-based Practices for Children with Special Needs and Disabilities*; Springer: New York, NY, USA. <https://doi.org/10.1007/978-1-4939-1483-8>
- IEA (2019). PIRLS 2021 Survey Operations Procedures Unit 2B – Sampling Classes and Field Test Administration [unpublished internal document]. IEA.
- IEA (2020). digitalPIRLS 2021 School Coordinator Manual [unpublished internal document]. IEA.

- IEA (2021a). ICCS 2022 Survey Operation Procedures Unit 4 – Data Collection and Quality Monitoring Procedures – Main Survey [unpublished internal document]. IEA.
- IEA (2021b). ICCS 2022 School Coordinator Manual – Main Survey [unpublished internal document]. IEA.
- IEA (2021c). TIMSS 2023 Survey Operations Procedures Unit 2B – Sampling Classes and Field Test Administration [unpublished internal document]. IEA.
- IEA (2022a). ICILS 2023 Survey Operation Procedures Unit 4 – Data Collection and Quality Monitoring Procedures – Main Survey [unpublished internal document]. IEA.
- IEA (2022b). ICILS 2023 School Coordinator Manual – Main Survey [unpublished internal document]. IEA.
- IEA (2022c). TIMSS 2023 School Coordinator Manual [unpublished internal document]. IEA.
- Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. §§ 1400–1482 (2004).
- Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interactions between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practice*, 27(3), 3–16.
<https://doi.org/10.1111/j.1745-3992.2008.00124.x>
- Kjeldsen, C. C., & Strietholt, R. (2024). Exclusions and non-response in contemporary international large-scale studies. *Educational Assessment, Evaluation and Accountability*, 36(3), 295–299.
- Klieme, E. (2012). The role of large-scale assessments in research on educational effectiveness and school development. In R. Strietholt, W. Bos, J.-E. Gustafsson, & M. Rosén (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 115–147). Springer Netherlands.
- LaRoche, S., Joncas, M., & Foy, P. (2020). Sample design in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 3.1–3.33). TIMSS & PIRLS International Study Center, Boston College.
<https://timssandpirls.bc.edu/timss2019/methods/chapter-3.html>

- Lovett, B. J., & Lewandowski, L. J. (2015). *Testing accommodations for students with disabilities: Research-based practice*. American Psychological Association. <https://doi.org/10.1037/14468-000>
- Martin, M. O., Fraillon, J., & Sibberns, H. (Eds.). (2025). IEA's technical standards for international large-scale assessment. International Association for the Evaluation of Educational Achievement. <https://www.iea.nl/sites/default/files/2025-07/IEAs-Technical-Standards.pdf>
- Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. TIMSS & PIRLS International Study Center, Boston College.
- Mayring, P. (2021). *Qualitative content analysis: A step-by-step guide*. SAGE Publications.
- Meinck, S. (2020). Sampling, weighting, and variance estimation. In H. Wagemaker (Ed.), *Reliability and validity of international large-scale assessment* (pp. 113–130). Springer. https://doi.org/10.1007/978-3-030-53081-5_7
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.
- Mezzanotte, C. (2023). *Indicators of inclusion in education: A framework for analysis* (OECD Education Working Papers No. 300). OECD Publishing.
- Musu, L. & Hartmanová, K. (2024). Quality Assurance in ICCS 2022. In W. Schulz, T. Friedman, & J. Fraillon (Eds.), *ICCS 2022 Technical Report* (pp. 95-116).
- National Center for Education Statistics. (1996). *Increasing the inclusion of students with disabilities and limited English proficient students in NAEP* (NCES 96-894). U.S. Department of Education. <https://nces.ed.gov/pubs96/web/96894.asp>
- National Center for Education Statistics. (2022). *NAEP accommodations policies*. U.S. Department of Education.
- OECD. (2012). *OECD child well-being module. CX3.1 Special educational needs (SEN)*. OECD Publishing.
- Quenemoen, R. F., & Thurlow, M. L. (2019). *Students with disabilities in educational policy, practice, and professional judgment* (NCEO Report 413). University of Minnesota.

- Schuelka, M. J. (2013). Excluding students with disabilities from the culture of achievement: The case of TIMSS, PIRLS, and PISA. *Journal of Education Policy*, 28(2), 216–230.
<https://doi.org/10.1080/02680939.2012.708789>
- Snozzi, R., Müller, C. M., & Zurbriggen, C. L. (2025). School placement trajectories of students with special educational needs—A longitudinal analysis of administrative data. *Journal of Research in Special Educational Needs*, 25(1), 103–117. <https://doi.org/10.1111/1471-3802.12714>
- Strietholt, R., & Johansson, S. (2023). Challenges for the design of international assessments: Sampling, measurement, and causality. *On Education. Journal for Research and Debate*, 6(18). https://doi.org/10.17899/on_ed.2023.18.2
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (Synthesis Report 44). University of Minnesota.
- Thurlow, M. L. (2021). *An updated state guide to universally designed assessments* (NCEO Report 431). University of Minnesota.
- Thurlow, M. L., Quenemoen, R. F., Thompson, S. J., & Lehr, C. A. (2001). *Principles and characteristics of inclusive assessment and accountability systems* (Synthesis Report 40). University of Minnesota.
- United Nations. (2006). *Convention on the Rights of Persons with Disabilities*.
<https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*.
<https://sdgs.un.org/2030agenda>
- UNESCO. (1994). *The Salamanca statement and framework for action on special needs education*.
<https://unesdoc.unesco.org/ark:/48223/pf0000098427>
- UNESCO. (2009). *Policy guidelines on inclusion in education*. UNESCO.
- UNESCO. (2015). *Education 2030: Incheon declaration and framework for action*.
<https://unesdoc.unesco.org/ark:/48223/pf0000245656>
- UNESCO. (2017). *A guide for ensuring inclusion and equity in education*. UNESCO.
- UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education—All means all*.
<https://unesdoc.unesco.org/ark:/48223/pf0000373718>

UNICEF. (2022). *Module on child functioning: Questionnaires*.

<https://data.unicef.org/resources/module-child-functioning/>

UNESCO. (2025). *Global education monitoring report 2025*.

https://unesdoc.unesco.org/ark:/48223/pf0000397618_eng

Voisin, C., Guez, A., & Piacentini, M. (2025). Improving accessibility in PISA: Guidelines for revising and designing PISA items. OECD Education Working Papers, (333), 0_1-27.

<https://dx.doi.org/10.1787/aa90f707-en>

Wagemaker, H. (Ed.). (2020). *Reliability and validity of international large-scale assessment* (IEA Research for Education, Vol. 10). Springer. <https://doi.org/10.1007/978-3-030-53081-5>

World Bank. (2023). Using Principles of Universal Design for Assessment (UDA) Design Accessible Learning Assessments - Toolkit (English). World Bank Group.

<http://documents.worldbank.org/curated/>

[en/099100223135027487/P1713680054b4908a0be920cbeb9acd2b2e](http://documents.worldbank.org/curated/en/099100223135027487/P1713680054b4908a0be920cbeb9acd2b2e)

World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. WHO.

Acknowledgements

We gratefully acknowledge Carmen Zahn and Luana Maria Gaiser Furtado for their assistance in coding and categorizing the open-ended survey responses. Their careful work contributed to the qualitative analysis presented in this paper.

Funding

This research was supported by the IEA Research and Development Fund.

Competing Interests

The authors declare no competing interests.

Author contributions

Ana Maria Mejia-Rodriguez, Nurullah Eryilmaz, Olga Kunina-Habenicht, and Rolf Strietholt all contributed to conceptualization, methodology, interpretation of results, writing — original draft, writing — review and editing, and visualization. Ana Maria Mejia-Rodriguez and Nurullah Eryilmaz were responsible for data collection and formal analysis.

Supplementary Material A: Detailed Methods

Overview

This study employs an international comparative approach to examine exclusion and accommodation practices in educational assessments conducted by the International Association for the Evaluation of Educational Achievement (IEA). We contend that systematic cross-national comparison can establish evidence on the current state of practice and, crucially, reveal the range of international variation in how students with difficulties are treated in assessment contexts. Documenting this variation is valuable because it makes visible what is possible: countries differ substantially in the extent to which they include and accommodate students, and these differences can inform efforts to expand participation.

Our empirical approach draws on a survey of national research coordinators who are responsible for the national implementation of IEA international assessments in their respective countries. National research coordinators have a unique advantage: they oversee sampling, apply exclusion criteria, and coordinate accommodations, making them the most knowledgeable informants on how international assessment guidelines are translated into national practice. Targeting national research coordinators also ensures a degree of comparability across countries, as all respondents operate within the shared procedural framework established by the IEA.

The survey instrument combines a comprehensive classification framework for student difficulties with a systematic inventory of accommodation practices. Items use both closed and open-ended response formats, enabling a mixed-methods analytic strategy. Specifically, we employ descriptive quantitative analyses to map patterns of inclusion, exclusion, and accommodation across countries and difficulty types, complemented by qualitative content analysis of open-ended responses to capture practices and distinctions that standardized items may not fully represent.

Conceptual Framework: Classifying Students by Functional Difficulties

A key conceptual decision in this study is the use of a functional, needs-based classification of students rather than diagnostic categories or umbrella terms such as “special educational needs” or “disability.” We refer to students with *difficulties* — functional limitations that may affect their ability to participate in an assessment. This choice is grounded in the biopsychosocial model underlying the

International Classification of Functioning, Disability and Health (ICF; WHO, 2001), which conceptualizes disability as the product of an interaction between a person's condition and the demands of their environment rather than as a fixed individual attribute. In an assessment context, what matters is not a student's diagnosis but the specific barriers they face when sitting a test.

A functional framing also supports cross-national comparability, since diagnostic categories and SEN definitions vary substantially across countries (Brussino, 2020; OECD, 2012), whereas functional difficulties can be described in terms that are consistent across education systems. Finally, classifying by difficulty rather than by diagnosis enables a more direct mapping to accommodations, since different diagnoses can produce similar functional barriers and the same diagnosis may manifest differently across individuals (Guez et al., 2024). Our approach builds on the needs-based classification proposed in the OECD's mapping study on accommodations in PISA (Guez et al., 2024) and is consistent with the UNICEF/Washington Group Module on Child Functioning and Disability (UNICEF, 2022).

Instrument Development

Data were collected through a purpose-designed online questionnaire administered to national research coordinators. The instrument was informed by prior research on exclusions and accommodations in ILSAs (Schuelka, 2013; Guez et al., 2024), by technical documentation from IEA studies (e.g., LaRoche et al., 2020; Almaskut et al., 2023; Cortes & Atasever, 2024), and by the OECD's mapping study on accommodations for students with SEN in PISA (Guez et al., 2024).

Classification of Functional Difficulties

The core of the instrument is organized around two classification frameworks developed for this study. The first is a typology of 17 functional difficulties¹ that may affect a student's ability to sit an assessment. These span six broad domains: intellectual disabilities; physical and sensory impairments (including moderate and severe difficulties seeing, difficulties hearing, severe mobility

¹ Intellectual disabilities, moderate difficulties seeing, severe difficulties seeing, difficulties hearing, minor mobility impairments, severe mobility impairments, fine motor/coordination difficulties, difficulties producing speech, difficulties understanding written text, difficulties with mathematics, difficulties controlling behavior, difficulties focusing, emotional difficulties, chronic illness/health conditions, limited language proficiency, severe/multiple disabilities, and difficulties with spoken language.

impairments, and fine motor difficulties); specific learning difficulties (including difficulties with written text, mathematics, and speech); emotional and behavioral challenges (including difficulties controlling behavior, focusing, and emotional and mental health difficulties); chronic health conditions; and limited proficiency in the language of assessment.

This classification was derived from three sources. First, we drew on the OECD mapping study (Guez et al., 2024), which proposed a needs-based categorization of barriers to assessment participation in PISA. Second, we aligned our categories with the ICF (WHO, 2001) and the UNICEF/Washington Group Module on Child Functioning and Disability (UNICEF, 2022) to ensure consistency with international standards. Third, we conducted an empirical analysis of 358 SEN category entries reported by national research coordinators in internal IEA sampling forms across ICCS 2022, ICILS 2023, and TIMSS 2023. These entries included 168 unique terms, many of which referred to the same underlying condition using different labels (e.g., visual impairment appearing as “vision,” “blind,” “low vision,” “visually disabled,” among others). All terms were systematically reviewed and consolidated into the 17 difficulty categories used in the survey.

Inventory of Assessment Accommodations

The second framework is a comprehensive inventory of 39 assessment accommodations², organized into five domains: presentation (14 accommodations, e.g., large print materials, Braille, sign language interpretation, screen readers, flexible digital interface), response (10, e.g., use of a scribe, speech-to-text technology, specialized keyboards, calculators), setting (7, e.g., individual testing, small-group testing, noise buffers, preferential seating), timing and scheduling (5, e.g.,

² Adjustments to print materials, adjustments to digital interface, magnification devices, adapted texts, simplified supports, Braille test materials, screen reader software/text-to-speech software, human reader, human translator, text translation (printed or digital text translation into another language), recorded instructions, sign language interpretation (live or recorded), cued speech (use of hand cues to show speech sounds), audio amplification devices (e.g., FM systems), human writer/scribe, speech-to-text technology (voice recognition software), word processor (typing responses on a computer), Braille writer/keyboard, calculators (basic, specialized, or four-function calculator use), spelling and grammar check (digital tools), highlighters (digital or paper-based), specific keyboard or mouse (adaptive input devices), specific writing materials (e.g., pencil grips, slant boards, special pens), assistive technology (e.g., eye-control, breath-control), small group testing, individual (one-on-one) testing, preferential seating, wheelchair-accessible testing locations, noise buffers (e.g., earplugs, noise-canceling headphones), changes to room conditions (e.g., lighting, temperature, furniture), minimal visual/auditory distractions, multiple-day testing, flexible scheduling, extended time, additional or extended breaks, shorter test, permission for food/drink during testing, permission for medication or medical devices use during testing, and monitoring by medical personnel.

extended time, additional breaks, flexible scheduling, multiple-day testing), and medical assistance (3, e.g., permission for medication use, blood sugar testing, and monitoring by medical personnel during testing).

These categories were drawn from accommodation taxonomies used in NAEP (NCES, 2022), TIMSS/PIRLS (IEA, 2021), and the College Board Accommodations and Supports Handbook (College Board, 2025), and reflect the principle that accommodations should remove construct-irrelevant barriers without altering the construct being measured (Fuchs & Fuchs, 1999; Lovett & Lewandowski, 2015).

Questionnaire Structure

The central section of the questionnaire asked national research coordinators to indicate, for each of the 17 difficulty types, which of the 39 accommodations their country provides. To ensure completeness and consistency, all 663 possible difficulty–accommodation combinations were presented, including combinations that may not be practically relevant (e.g., magnification devices for students with hearing difficulties). This design choice was made deliberately to avoid imposing assumptions about which accommodations countries might or might not offer for particular difficulties, and to allow unexpected practices to surface in the data.

Beyond accommodations, the questionnaire addressed several additional dimensions. National research coordinators reported on the educational settings in which students with each type of difficulty are typically schooled—whether in regular classrooms, special classes, or special schools. They also indicated the proportion of students with each difficulty type who are included in or excluded from IEA assessments and identified who is involved in making inclusion or exclusion decisions. Finally, an open-ended section asked national research coordinators what resources or support they would need to include more students with difficulties in future assessment cycles.

Sample

To enable meaningful cross-national comparison, this study focuses on international large-scale assessments conducted by the IEA. A key advantage of these assessments is their high degree of international standardization: participating countries draw large probability-based samples from defined target populations and administer the same instruments under comparable conditions. This

standardization makes cross-national differences in exclusion and accommodation practices interpretable, because variation in how students with difficulties are treated occurs against a shared procedural backdrop rather than reflecting fundamentally different assessment designs.

We surveyed the national research coordinators responsible for implementing these assessments in their respective countries. National research coordinators were invited to complete the survey for each IEA study in which their country participated in the most recent cycle. For PIRLS, national research coordinators were surveyed during the preparation phase of the 2026 cycle and reported on the field trial stage, rather than the final study implementation; all other assessments (TIMSS 2023, ICCS 2022, ICILS 2023) had completed data collection at the time of the survey.

In total, 234 national research coordinators were invited across five assessments: TIMSS 2023 Grade 4 ($n = 65$), TIMSS 2023 Grade 8 ($n = 47$), PIRLS 2026 ($n = 63$), ICCS 2022 ($n = 24$), and ICILS 2023 ($n = 35$). As many education systems participate in multiple IEA studies and some national research coordinators fulfil this role for more than one study, these 234 invitations translated into a coverage of 85 countries. A total of 135 responses were received (response rate: 58%), with the following breakdown by study: TIMSS 2023 Grade 4 ($n = 37$), TIMSS 2023 Grade 8 ($n = 20$), PIRLS 2026 ($n = 41$), ICCS 2022 ($n = 17$), and ICILS 2023 ($n = 20$). These responses covered a total of 63 countries. In one case (for the four education systems within the United Kingdom participating in PIRLS), a single response was provided consolidating information for the four systems. This response was treated as one response unit. In a few other cases, responses were duplicated across systems for different studies at the request of national research coordinators, as the responses would be identical between studies.

Because many countries participate in multiple IEA studies, the unit of analysis is the country–study combination rather than the country. This means that a single country can contribute up to five observations, one per assessment, reflecting the fact that exclusion and accommodation practices may differ across studies even within the same national context.

Geographic and Regional Coverage

The 63 participating countries and subnational entities span a broad range of world regions and socioeconomic contexts. The sample includes substantial representation from Europe, the Middle

East and North Africa, and the Asia-Pacific region, and the Americas. The full list of participating countries was: Albania, Armenia, Australia, Austria, Azerbaijan, Belgium (Flemish Community), Belgium (French Community), Bosnia and Herzegovina, Brazil, Canada, Canada (Quebec), Côte d'Ivoire, Croatia, Cyprus, Czech Republic, Denmark, United Kingdom (England, Scotland, Northern Ireland, and Wales), Finland, France, Georgia, Germany, Germany (North Rhine-Westphalia), Greece, Hong Kong, Hungary, Iran, Iraq, Ireland, Israel, Italy, Jordan, Kazakhstan, Kosovo, Latvia, Lithuania, Luxembourg, Macao SAR, Macedonia, Malaysia, Malta, Montenegro, Morocco, Netherlands, New Zealand, Norway, Palestine, Poland, Portugal, Qatar, Republic of Tajikistan, Romania, Saudi Arabia, Serbia, Slovakia, Slovenia, Spain, Sweden, Taiwan, Türkiye, United Arab Emirates, United Arab Emirates (Dubai), Uruguay, Uzbekistan.

Few education systems participated as subnational entities reflecting the decentralized governance of education in these contexts (e.g., Belgium's Flemish and French Communities, the Canadian province of Quebec). While the sample covers considerable geographic and economic diversity, it should be noted that European systems are overrepresented relative to countries in Sub-Saharan Africa, Latin America, and South and Southeast Asia, which limits the generalizability of findings to these regions.

Analysis

We employed a mixed-methods approach combining quantitative and qualitative analysis. For closed-ended items — including the proportion of students included or excluded by difficulty type and the use of specific accommodations — we computed frequency counts and cross-tabulations to map patterns across countries, assessments, and difficulty categories. For open-ended responses — such as descriptions of decision-making processes and the resources national research coordinators identified as necessary for greater inclusion — we applied qualitative content analysis to identify recurring themes and practices that standardized items may not fully capture.

All analyses were conducted at the country–study level, meaning that each combination of country and assessment constitutes one observation. Because many countries participate in more than one IEA study, this design could lead countries with multiple responses to be overrepresented in aggregate results. To address this, we applied inverse frequency weights: each country's responses

were weighted by the reciprocal of the number of assessments for which it provided data. For instance, a country providing responses for two assessments received a weight of 0.5 per response, while a country providing responses for three received a weight of approximately 0.33. This ensures that each country contributes equally to aggregate statistics regardless of how many assessments it covers.

The responses to two open-ended questions were analyzed independently by two raters using a qualitative content analysis approach as outlined by Mayring (2021). Following Mayring's structuring content analysis, a deductive–inductive procedure for category development was applied. The initial coding categories were derived deductively from the theoretical framework and the research questions. During the coding process, the category system was iteratively reviewed and refined, and additional categories were developed inductively to capture recurrent themes emerging from the data. All responses were coded according to this structured category system. In cases of disagreement, the raters discussed the respective coding decisions until consensus was reached.