A peer reviewed, open-access electronic journal: ISSN 1531-7714

Giving voice to stakeholders on fair and valid use of accommodations and universal tools in digital standardized tests

Sabine Dierick, *Hasselt University*
Fauve De Backer, *Ghent University*
Katrien Struyven, *Hasselt University*

Abstract: This study explores stakeholders' perspectives on the use of accommodations and universal tools in standardized digital assessments, focusing on (1) students with special educational needs (SEN) and (2) all students. Using data from 20 focus groups, representing educators, policy staff, parents, psychometricians, and legal experts (n = 182), we conducted thematic analysis to identify key considerations for inclusive assessment. The findings highlight persistent tensions between fairness, construct validity, and comparability, alongside broad support for Universal Design for Assessment (UDA) and the provision of individual accommodations. The study contributes to the ongoing discourse on establishing inclusive testing environments in large-scale assessments that are fair to all students, while upholding the principles of validity, reliability, and equity. Practical recommendations and policy scenarios are provided to guide the development of inclusive and valid large-scale testing systems.

Keywords: Standardized testing, Large-scale testing, Inclusive test environment, Universal design of assessment, Digital tools, Accommodations, Stakeholders, Students with Special Educational Needs

Introduction

Inclusion has become central to educational policy, particularly in assessment systems aimed at measuring students' knowledge and skills. Accommodations and accessibility features are essential for enabling fair participation in large-scale assessments, particularly for students with special educational needs (SEN), including those with disabilities or limited language proficiency (Thurlow & Kopriva, 2015). Accommodations are defined as changes to test format or administration that maintain the intended construct and yield comparable results to those of students not using such accommodations (AERA et al., 2014; Thurlow et al., 2005). These may include adjustments to content presentation, test setting, response mode, timing, or the use of additional resources or devices during the test (Sireci & O'Riordan, 2020;

Thurlow, 2005). Unlike modifications, accommodations do not alter what is being measured (Lazarus et al., 2009).

The goal of accommodations is to provide all students with a fair opportunity to take a particular test by eliminating barriers that would impede their performance. Fairness, as defined in the testing standards, requires responsiveness to individual differences so that test scores provide valid interpretations for intended uses (AERA et al., 2014). To ensure meaningful feedback, assessments must allow all learners to express their knowledge appropriately (Dolan & Hall, 2001). However, in large-scale testing, not all accommodations are permitted, even for students with SEN. This creates a tension between the principle of standardization and the goal of fairness. Furthermore, some accommodations originally designed for students with SEN may also benefit the broader student population.

As digital assessment platforms have become more prevalent, the idea of providing certain tools as universal tools has gained prominence. Universal tools are digital tools that are embedded and available to all students by default, regardless of their learning status. Examples include zoom functions, digital highlighters, spell-checkers, calculators, and text-to-speech with regard to instructions. These tools are designed to improve accessibility without altering the construct being evaluated. To avoid terminological ambiguity in digital testing environments, this article uses the term "digital support tools" to refer to all ICT-based support tools available during testing. This includes universal tools for all students, as well as accommodations for students with SEN. Digital support tools present a unique opportunity to improve accessibility provided they do not alter the intended construct (International Test Commission & ATP, 2022; Smarter Balanced Assessment Consortium, 2014). Accessibility itself refers to ensuring that all test takers can fully demonstrate their standing on the construct being assessed (AERA et al., 2014).

Much of the literature on Universal Design of Assessment (UDA) emphasizes the need for thorough policy planning to enable the inclusive design of large-scale, valid, and reliable tests (Hanna, 2005). A sound testing policy requires the support of all stakeholders involved in the development and use of standardized testing and in determining the educational and policy implications. To ensure fairness in standardized testing, it is crucial that the test-taking policy governing accommodations or the use of universal tools is formulated as concretely as possible, and applied in the most standardized manner. This is vital to avoid undesirable discrepancies between schools.

To gain a comprehensive understanding of fair and valid accommodation and universal tool use in digital large-scale testing, this study examines the perspectives of various stakeholders on this topic. Specifically, the study will focus on the following key stakeholder groups: (1) care coordinators, (2) subject-specific teachers, (3) educational policy staff, (4) parents of students with SEN, (5) psychometric experts, and (6) legal representatives.

In the following sections, we first explore accommodations to minimize construct-irrelevant variance, then consider the balance between fairness and standardization. Next, we examine digital testing as an opportunity for accessibility, followed by a discussion of stakeholder roles in shaping inclusive assessment practices.

Testing accommodations for students with specific educational needs

A primary tenet of standardized testing is strict adherence to standard assessment procedures when administering tests. This principle aims to provide all students with an equal opportunity to perform well on the test. For performance to accurately reflect a student's true ability, it is essential to minimize "construct-irrelevant variance" (Messick, 1989).

Construct-irrelevant variance occurs when differences in student performance on tests arise not only from the construct or underlying ability being measured but also from other student differentiating

characteristics. For instance, students with dyslexia who possess the same mathematical ability as their peers may perform poorly on a mathematics test due to difficulties in reading the instructions. Differences in reading ability—characteristics separate from the construct being measured—should not influence mathematics achievement. Therefore, accommodations should only be permitted to avoid construct-irrelevant variance. Accommodations themselves should not interfere with the construct being measured (NAEP, 2014) but should aim solely to eliminate differences between students caused by contextual barriers.

Determining which accommodations are best suited for which learners is not a straightforward process (De Backer et al., 2023) and should also be considered in relation to ensuring the construct validity of the test. Thus, the feasibility of allowing accommodations depends on the construct being measured. Permitting accommodations should be preceded by an assessment of their potential impact on construct validity.

Content-related accommodations

Accommodations that contain information or resources are referred to in this study as "*content-related accommodations*". Examples include calculators, text-to-speech software (TTS), dictionaries, or formularies. These accommodations may affect the construct validity of a test. For instance, TTS can enhance access for students with reading difficulties but may also affect the validity of tests designed to measure reading comprehension. The use of a calculator is essential for students with dyscalculia, as it alleviates memory load; however, if the test aims to measure a student's numeracy, allowing a calculator alters the construct being measured.

Content-related accommodations modify standardized procedures, which are intended to be uniform. Nevertheless, without such accommodations, many examinees with disabilities could not be adequately assessed. So, the principles of standardization and fairness can sometimes conflict, a situation which Sireci and O'Riordan (2020) describe as a dilemma. Providing an accommodation based on fairness may change the construct measured, and in some cases this change may make testing easier for those who receive an accommodation. The effect of accommodations on the constructs measured by a test is therefore a critical validity issue that directly affects score comparability across standard and accommodated tests (Sireci & O'Riordan, 2020).

From a measurement perspective, empirical research on the impact of accommodations on test scores is the best way to support the provision of such accommodations to groups of students. Some researchers consider an accommodation to be fair only if it raises the test scores of students who need the accommodation, while leaving the test scores of students who do not need it unchanged (Buzick & Stone, 2014; Zuriff, 2000). Thus, they expect the accommodation to produce an *interaction effect,* i.e. a differential effect for certain learners (groups), rather than a main effect. If an accommodation not only helps to reduce contextual factors for specific groups of students, but also contributes to the performance of regular students, then the accommodation affects the construct to be measured.

The *differential boost hypothesis* is similar but represents a more realistic depiction of the effectiveness of accommodations by relaxing the hypothesis that students without SEN will not have score gains in the accommodation condition. According to the differential boost hypothesis, if an accommodation is effective, the gains for SEN will be greater than the gains observed for non-SEN (Cahalan-Laitusis, 2007; Kettler, 2012; Sireci, Scarpati & Li, 2005; Sireci & O'Riordan, 2020). As with the interaction hypothesis, differential boost is evaluated using experimental designs where one factor is the student group and the other factor is the test administration condition.

However, empirical research on the (differential) effectiveness of accommodations is very intensive and therefore rather scarce. More research on accommodations is available for students with disabilities (SWD) than for multilingual students (MS), and the accommodations for SWD are often used for MS without appropriate validation (Li & Suen, 2012). Generalizing the findings of accommodations for SWD to MS

should not be done lightly, since both groups have different characteristics and different assessment needs (Li & Suen, 2012; Solano-Flores, 2016).

Furthermore, the results of such studies cannot simply be generalized to other testing contexts. Nor is it feasible to validate every accommodation with an effectiveness study with different groups of students before implementing it. As a result, developing a policy for inclusive large-scale testing is quite complex (Bolt & Thurlow, 2004).

International exploration shows that the use of content-related accommodations such as calculators or formularies, is either prohibited as in the Netherlands (Bureau ICE, 2019), or its use is restricted in large-scale tests, as seen in Australia, England, or the USA (Australian Curriculum, Assessment and Reporting Authority [ACARA]; Joint Council for Qualifications, 2021; National Assessment Governing Board, 2014; National Center for Education Statistics, 2019, 2021).

Additionally, the use of TTS is not always permitted during reading tests, even if the student is accustomed to using it in class. If accommodations are not allowed, it is possible under certain conditions to exempt students with severe disabilities, such as is the case in England (Joint Council for Qualifications, 2021) or multiple disabilities from one or more test items, as in Australia (ACARA, 2021), or to exclude the results of students with accommodations for those test items from the analysis, as in the USA (National Assessment Governing Board, 2014; National Center for Education Statistics, 2019, 2021).

Digital testing as an opportunity for accessibility: from an accommodation for students with SEN to a universal tool

The use of digitally delivered tests has expanded opportunities to enhance the accessibility of test items and interfaces (International Test Commission and Association of Test Publishers, 2022).

One way to promote accessibility is to permit accommodations as digital support tools for students likely to encounter construct-irrelevant barriers during testing, including, but not limited to, those with disabilities or second language learners. The test platform can automatically provide (embedded) digital support tools allowed on a test, such as calculators, dictionaries, or a TTS button. This could even be specified per item within the test to monitor construct validity. It requires, at a minimum, that students gain experience with the range of tools in classroom practice or that the digital support tools are so straightforward to use that they enhance accessibility without altering the construct intended to be measured by the assessment (Sireci & O'Riordan, 2020). This argument aligns with studies on the effectiveness of accommodations, which demonstrate their potential benefits for all learners (Johnstone, 2003; Laitusis et al., 2012).

A potential drawback is that some students might be disadvantaged during testing due to distractions relating to the digital support tools that meet the access needs of others. Problems can arise when tests are presented through multiple modalities, e.g., written test questions supported by auditory clarification through TTS. Although research results are not conclusive, several studies on the use of dual modalities in reading comprehension suggest that students perform worse when tests and learning materials are presented through multiple modes. It is hypothesized that simultaneous input from two modes increases cognitive load for learners, thereby diminishing learning performance (Plass et al., 2010).

The goal of an accommodation is to assist those students who need it without affecting other students (Shaftel et al., 2006). It is also questionable whether making digital support tools available can meet all requests for accommodations for students with SEN. In addition to digital support tools, students with SEN may require local accommodations, such as taking the test in a separate room or using an accustomed formulary. It may also be determined that accommodations should only be offered to specific groups of students who can clearly demonstrate their need for such accommodations based on their SEN. Accommodations should then be provided on an individual basis.

The aforementioned arguments illustrate that the design of an inclusive standardized test that accommodates students with SEN and provides digital support tools for all students in a digital test environment is a complex endeavor. Furthermore, depending on the expertise of a stakeholder, some arguments may of greater importance than other arguments.

Giving voice to stakeholders about the use of accommodations and universal tools in large-scale digital testing

The primary goal of large-scale testing is to provide a fair and valid assessment of students' proficiency. Accommodations are necessary to give students who face barriers the opportunity to demonstrate their abilities, including students with disabilities and with limited language proficiency.

As discussed in the literature, the use of content-related accommodations in these tests is not straightforward, as it alters standardized procedures which are designed to be uniform. However, without such accommodations, many students with SEN would not be adequately assessed. International legislation, such as the Salamanca Declaration (UNESCO, 1995), also indicates that students with SEN are entitled to accommodations. Therefore, educational and legal arguments for allowing accommodations based on fairness sometimes appear to conflict with the principles of standardization in large-scale testing.

Moreover, some accommodations are also beneficial for students without SEN. Digital testing offers opportunities for accessibility for all students. The question of how accommodations can be offered as digital support tools can generate debate, particularly regarding whether certain tools can be used in large-scale tests, which also touches on teachers' didactic views and principles.

This apparent contradiction regarding the use of accommodations and universal tools in large-scale digital tests underscores the value of hearing from all stakeholders involved in their development and use, as well as discovering their educational and policy perspectives on this issue. This may be a priority for psychometricians, as well as for parents, teachers, care coordinators, subject-specific teachers, legal representatives, and educational policy staff such as school management, pedagogical supervisors, and educational inspectors. In addition to legal and educational perspectives, pedagogical supervisors—who support the implementation of education and care policies—tend to emphasize feasibility and alignment with broader inclusion goals. Inspectors, responsible for evaluating educational policy, are more likely to focus on whether digital tools support learning and meet psychometric standards.

As has been pointed out above, empirical research on the differential effectiveness of accommodations is scarce, and results cannot be generalized to other testing contexts. In addition, it is not feasible to validate each accommodation with an effectiveness study on different groups of students prior to implementation. This study, therefore, explores the face validity of fair and valid accommodations and universal tool use according to different stakeholders: (1) care coordinators, (2) subject-specific teachers, (3) educational policy staff, (4) parents of students with SEN, (5) psychometric experts, and (6) legal representatives. Giving voice to the diverse stakeholders involved in large-scale testing provides a comprehensive understanding of this complex issue.

In the context of inclusive administration of large-scale tests, *two major points of discussion* will be further investigated in the present study. According to the different perspectives (i.e., care, subject specific teachers, policy, psychometric, legal, or as a parent) of stakeholders:

(1) When are accommodations for SEN allowed to be used in digital large-scale testing?

(2) How can accommodations be made available to all students as universal tools in digital large-scale testing?

Methodology

Research context

Flanders has a long tradition of standardized testing, organized by various educational umbrella organizations. Starting in Spring 2024, centralized standardized tests at the Flemish level was made compulsory for all students. These tests are to be administered digitally at the end of the second and third grades of primary education and the first and third grades of secondary education.

The primary goal of these tests is to provide a fair and valid assessment of students' proficiency in both mathematics and Dutch. Students who face barriers to demonstrating their ability, including those with disabilities and limited Dutch language skills (the official school language), require appropriate support. Flanders has a clear framework and legislation with regard to inclusive education, based on the Salamanca Declaration (UNESCO, 1995), which states that students with SEN are entitled to the use of accommodations. The provider of centralized tests is therefore obliged to provide them when administering the test.

In collaboration with different stakeholders the present study aims to develop a framework for all possible accommodations and universal tools to meet the needs of the full range of students expected to participate in these large-scale tests, including students with SEN and those with limited language skills in Dutch (see Figure 1). A distinction is made between tools that can be offered digitally, and those that need to be provided locally in the classroom (where the test takes place). The goal is to embed accommodations and universal tools within the digital platform.

Figure 1. Framework of Needed Test Accommodations and Universal Support in Large-Scale Centralized Testing in Flanders

Locally Provided Test Accommodations and Universal Support		
	Universal support	**Individual Test Accommodations**
☐ Timing		☐ supervised breaks ☐ spread in tests
☐ Answer		☐ writer (someone who registers answers from students) ☐ have answers orally reviewed by trusted person/teacher
☐ Environment material	☐ headphones ☐ earbuds	☐ custom furniture ☐ wiggle and fidget materials ☐ concentration screen (e.g., Study Buddy) ☐ chewing gum (focus) ☐ quiet room ☐ separate place in classroom ☐ low-stimulation environment (auditory, visual, tactile, etc.)
☐ Environment others		☐ support person or supervisor ☐ presence of familiar person or familiar object ☐ personal assistant
☐ Tools	☐ ruler/geodril ☐ notepad ☐ table chart ☐ formulary	☐ customized formulary/memorization card ☐ braille reader ☐ custom mouse/keyboard

	☐ inference table ☐ writing frame	☐ audio system that can be connected to FM system for hearing impaired students with cochlear implant/hearing aid ☐ custom geodrite/passer/... ☐ TV reading magnifier
☐ Presentation	☐ oral explanation	☐ braille ☐ interpreter Dutch - Flemish Sign Language
Digital Embedded Test Accommodations and Universal Support		
	Universal Support	**Individual test accommodations**
☐ Timing		☐ extra breaks ☐ spread in tests ☐ limiting number of exercises ☐ visualizing remaining time (e.g. time timer)
☐ Answer		☐ software to dictate answers (e.g. Dragon)
☐ Tools	☐ zoom/magnifier ☐ dictionary (with image) ☐ highlighting function ☐ digital notepad ☐ reading software for instructions ☐ spell check ☐ calculator ☐ formulary ☐ glossary	☐ simple calculator ☐ text to speech ☐ spell check ☐ word prediction ☐ NT2 dictionary (multilingual students) ☐ translation application (e.g. SayHi)
☐ Presentation	☐ easy navigation ☐ plain layout ☐ choice of color contrast ☐ no unnecessary images/details ☐ clearly readable font ☐ choice between fonts ☐ possibility to add comments to question ☐ possibility to skip question for now ☐ volume button for sound clips ☐ subtitling of video fragments ☐ short, clear and concrete instruction ☐ division of assignments into small subtasks ☐ clarify expectations around the answer ☐ clarify importance of the question ☐ instructional videos	☐ at media clips: interpreter Dutch - Flemish Sign Language in the corner ☐ statements in accessible format (e.g. braille, large print, on paper, etc.)

Participants and procedures

This qualitative study involved 20 focus groups with 182 participants from primary and secondary education, representing care coordinators, subject teachers, policy staff, parents, psychometricians, and legal experts (see Table 1).

Table 1. Participants in Focus Group Discussions by Stakeholder Perspective, Role, Region, and Education Level

Perspective	Function	Region	PE	SE	Other	Total
Care	Care Coördinators	Antwerp	10	3		13
		Ghent	8	3		11
		Hasselt	10	1		11
		Leuven	10	4		14
	Total Care		**38**	**11**		**49**
Teachers	Math Teachers	Antwerp		3		3
		Ghent		7		7
		Hasselt		3		3
		Leuven		3		3
	Total			***16***		***16***
	Dutch Teachers	Antwerp		3		3
		Ghent		4		4
		Hasselt		3		3
		Leuven		3		3
	Total			***13***		***13***
	Total Teachers			**29**		**29**
Policy	Pedagogical Supervisors	Antwerp	6	5		11
		Ghent	6	8		14
		Hasselt	4	6		10
		Leuven	4	5		9
	Total		***20***	***24***		***44***
	Members of Inspection	Ghent	2	1		3
		Hasselt	1	2		3
		Leuven	2	2		4
	Total		***5***	***5***		***10***
	School Management	Antwerp	2	4		6
		Ghent	7	3		10
		Hasselt	6	3		9
		Leuven	3	2		5
	Total		***18***	***12***		***30***
	Total Policy		**43**	**41**		**84**
Parents	Parents of Children with SEN				9	**9**
Psychometric	Researchers with Expertise in Assessment				7	**7**
Legal	Legal Experts in Education Laws				4	**4**
Grand total			**81**	**81**	**20**	**182**

Note. PE= Primary Education; SE= Secondary Education

Focus groups were conducted in the fall of 2021 across four cities (Hasselt, Leuven, Antwerp, and Ghent) and organized by stakeholder perspective to capture distinct viewpoints. Sampling aimed to ensure geographic and institutional diversity. A random selection of 10% of schools across networks and provinces was used to recruit care coordinators and subject teachers (mathematics and Dutch).

In secondary education in Flanders, subject-specific teachers are responsible for teaching Dutch and mathematics. These teachers have expertise in their subjects and understand how tools function during instruction and assessment. Meanwhile, the care coordinator plays a pivotal role in overseeing the

implementation of individual accommodations and universal tools at the student, classroom, and school levels. Therefore, we invited both subject-specific teachers and the care coordinator from secondary schools.

In primary education, however, one teacher typically instructs in all subjects, and the care coordinator plays a more prominent role in coordinating support and assigning tools. For this reason, we only invited the care coordinator from primary education. Half of the sampled schools were invited to nominate participants for the care (1) and teacher perspectives (2); the other half were invited to participate in the policy perspective (3).

With regard to the policy perspective, invitations were sent to principals, pedagogical supervisors, and inspection staff, with dissemination supported by the heads of the Flemish Inspectorate and pedagogical guidance services. School principals, together with their team, are responsible for the quality of education in their schools. While educational supervisors guide and support schools in their activities to ensure the quality of education, the Education Inspectorate has a more supervisory role in monitoring the quality of education in Flanders.

Parents (4) were recruited through representative SEN advocacy groups, and only parents of students with SEN were included. Psychometricians (5) were selected based on expertise in assessment, and legal experts (6) were invited from a university legal education department and a legal education organization.

All 20 focus groups were organized by perspective in order to clearly capture the perceptions of each stakeholder group separately. Table 1 shows the diverse representation of stakeholders across educational networks and provinces, as well as all existing interest groups for children with disabilities in Flanders.

During the focus group discussions, we worked with several real-life case studies involving students with SEN (see Figure 2). Only content-related accommodations were selected for discussion in the case studies, as allowing this type of accommodation presents a dilemma in centralized testing. The case of Berten concerns a student with SEN in mathematics, while the case of Thibaut involves a student with SEN in the Dutch language. Both students have received compensatory measures from the 4th year onward, including the use of a calculator, a formulary, TTS with a word processor, a spell-checker, and a dictionary. Berten was diagnosed with a learning disability, while Thibaut was not. Finally, the case of Susa involves a Polish girl who has been living in Flanders for 10 months, and whose teacher sought various strategies to assist her, such as using tools to overcome the language barrier, including TTS, word prediction, and an explanatory dictionary.

The *first research question* addresses when accommodations can be used for students with SEN to measure specific skills during centralized testing. Participants were asked to vote individually through a digital system, "wooclap", on whether they believe the accommodations in the case studies should never, sometimes, or always be used for the central tests for Dutch (reading, comprehension and writing skills) and mathematics (numeracy: operations with fractions). Following an anonymous vote, the participants were encouraged to briefly explain, clarify, and justify their answers. They were always asked to contextualize their responses in relation to those of other participants (Baert, 2012).

The *second research question* explores how accommodations can be made available to all students in terms of digital centralized testing. Participants were asked to indicate individually whether the accommodations discussed in the case study should be made available in the digital testing environment as universal tools for all students, or as accommodations specific for students with SEN.

Figure 2. Example of a Real-life Case Study of a Student with Specific Educational Needs

Case Study Berten

Berten has problems with math in class. These problems became clearly visible in the second grade of primary education. When the teacher, together with the care coordinator, takes a close look at his math skills, it appears that Berten is constantly counting on his fingers. He has difficulty automating basic knowledge. His short-term memory can only handle a limited load, so that knowledge is not sufficiently imprinted. Simple operations (4×3=12) are not available automatically. He must calculate them over and over again. This process is slow and the chance of error is high. As a result, a fluent mastery of the operations is not ingrained.

Berten can do insightful math and still make mistakes. The lack of automation demands too much of his memory. He often knows efficient computational strategies but does not use them spontaneously. He has difficulty remembering or inferring the intended strategies. This interferes with the generalization of what he has learned.

Starting in the second grade of primary education, Berten receives the following accommodations:
– Everything that is not mental arithmetic can be solved with a calculator (metric arithmetic, math stories, etc.).
– The customized formulary can always be used.
– Use flash cards for mental arithmetic
– Allow more time for tests.

Question 1: Assume that Berten is a student who takes the central test in mathematics at the end of the third grade of primary education.

Is Berten allowed to use the following tools in the arithmetic test (operations with fractions)?

1. A calculator
2. A customized formulary

Question 2: Assume that Berten is a student who takes the central test in mathematics at the end of the first grade of secondary education.

Is Berten allowed to use the following tools in the arithmetic test (operations with fractions)?

1. A calculator
2. A customized formulary

Data analysis

Data from the focus groups were processed using thematic analysis (Vaïsmoradi et al., 2013). The voting data were utilized to make statements within a stakeholder group and between groups. A table with the quantitative voting results has been included (see Figure 3).

Figure 3. Overview of the Quantitative Voting Results and the Arguments given for the Use of Different Accommodations and Universal Tools in Standardized Testing from the Different Perspectives on Research Question 1

Tool	Use Category	Summary of Arguments	Perspectives	N
Calculator	Always	Fair when assigned by decree; aligns with class use; removes memory load but preserves conceptual understanding; supports dyscalculia.	Parents Legal Policy Teachers Care Psych	9 (100%) 4 (100%) 63 (75%) (PE: 34; SE: 29) 21 (72%) (SE: 21) 37 (76%) (PE: 29; SE: 8)
Calculator	Sometimes	Not needed for simple operations solvable with strategies if more time is given.	Policy Care	21 (25%) (PE: 9; SE: 12) 12 (16%) (PE: 9; SE: 3)
Calculator	Never	Mental arithmetic constructs require calculator-free performance.	Teachers	8 (8%) (SE: 8)
Customized Formulary	Always	Fair when assigned by decree; consistent with class use; answers not on card.	Parents Legal Policy Care Psych	9 (100%) 4 (100%) 74 (88%) (PE: 41; SE: 33) 49 (100%) (PE: 38; SE:11) 7 (100%)
Customized Formulary	Never	May compromise comparability as content differs across schools.	Policy	10 (12%) (PE: 2; SE: 8)
Word Prediction	Always	Fair when assigned or used in class; appropriate for functional writing.	Parents Legal Policy Care Teachers Psych	9 (100%) 4 (100%) 84 (100%) (PE: 43; SE: 41) 44 (90%) (PE: 38; SE: 6) 20 (69%) (SE: 20) 7 (100%)
Word Prediction	Sometimes	Vocabulary is part of writing construct; prediction not suitable when measuring vocabulary.	Teachers Care	6 (21%) (SE: 6) 5 (10%) (SE: 5)
Word Prediction	Never	Can distract students or promote speed over accuracy; suggestions may be incorrect.	Teachers	3 (10%) (SE: 3)
Text-to-Speech	Always	Fair for SEN; decoding barriers hinder comprehension; aligns reading and listening; supports functional understanding.	Parents Legal Policy Care	9 (100%) 4 (100%) 84 (100%) (PE: 43; SE: 41) 49 (100%) (PE: 38; SE: 11)

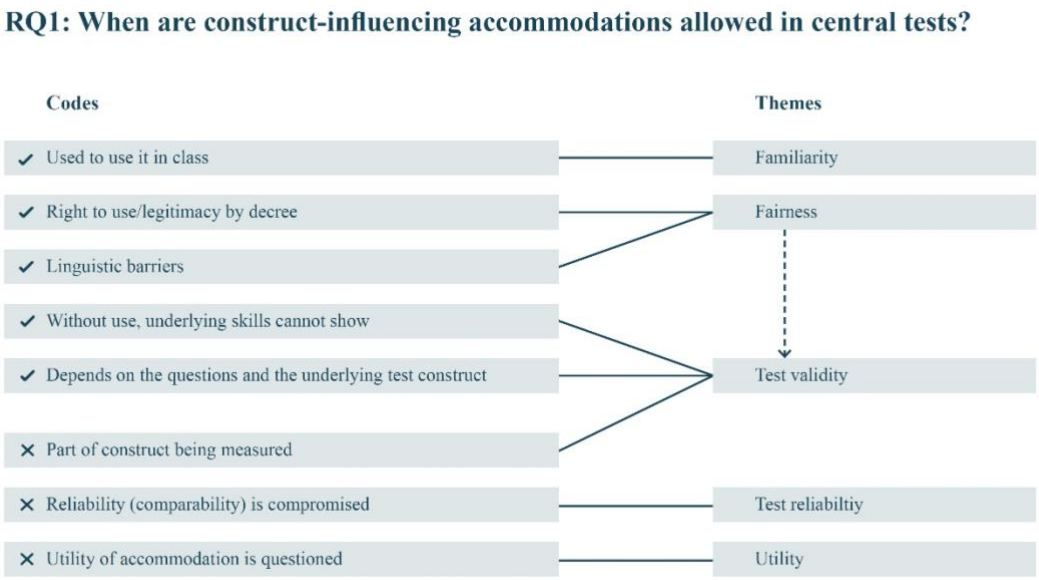
			Teachers Psych	29 (100%) (SE: 29) 7 (100%)
Dictionary	Always	Supports comprehension when words are unknown; aligns with functional reading.	Parents Policy Care Teachers Psych	9 (100%) 84 (100%) (PE: 42; SE: 41) 49 (100%) (PE: 38; SE: 11) 24 (83%) (SE: 24) 7 (100%)
Dictionary	Sometimes	Not allowed when meaning must be inferred from context.	Teachers	5 (17%)
Spell Check	Always	Spelling is supportive skill; students shouldn't be penalized when writing quality is measured.	Parents Policy Care Teachers Psych	9 (100%) 73 (87%) (PE: 36; SE: 37) 38 (75%) (PE: 31; SE: 7) 15 (52%) (SE: 15)
Spell Check	Sometimes	Not allowed when spelling rule is construct; allowed when meaning expression is goal.	Care Teachers	11 (22%) (PE: 7; SE: 4) 8 (28%) (SE: 8)
Spell Check	Never	Used when spelling is part of construct.	Teachers	6 (21%) (SE: 6)

Note. PE= Primary Education; SE= Secondary Education; Psych= Psychometric

To accurately represent the discussions during the focus groups, the conversations were transcribed verbatim from audio recordings. These transcriptions served as the foundation for the analysis. A data-driven approach was adopted for the coding methodology (Glaser & Strauss, 1976). The coding was derived from the empirical data collected during the focus group interviews (i.e., "bottom-up coding"). Senior researchers with expertise in inclusive education—familiar with international frameworks (e.g. Ainscow, Booth, & Dyson, 2006; Florian, 2014; UNESCO, 2020) and Flemish care policy—conducted the coding. In Flanders, care policy is based on a continuum of care, extending inclusive support across the different tiers of a multi-level support system (Flemish Ministry of Education and Training, 2017).

To address the potential for bias, the research team discussed diverging interpretations throughout the coding process and engaged in consensus-building to enhance the credibility of the findings. During coding, an analysis framework was employed where each category formed a row. Participants' statements were included verbatim (as quotes) in the analysis framework. The columns represented different perspectives, making the differences and similarities between and within the perspectives clear. Recurring patterns emerged in the various interviews, signaling data saturation (Fusch & Ness, 2015). The result of this iterative coding process was the delineation of themes (see Figure 4). The identified themes describe important patterns within the coding, and effectively represent the content of the qualitative data collected. "Literal" statements made by participants are presented for illustrative purposes in the form of quotes in italics in the results section.

Figure 4. Coding Tree Thematic Analysis



The results of the thematic analysis were shared with all participants in the study, asking for confirmation, nuance, or further additions as a form of "member check". All participants indicated that the text accurately reflected the content of the focus discussions. Two participants requested certain quotes to be complemented or nuanced. One participant noted the importance of consistently mentioning that the use of a tool can be "always, sometimes, or never" allowed in the specific case discussed. A final comment was made regarding the term "simple calculator", asking whether this could be specified as "a calculator with only the main operations that respects the order of operations".

Results

The following is a discussion of the results of this study in relation to the research questions that were set.

According to the different perspectives when are accommodations for students with special educational needs allowed to be used?

Several themes emerged from the data set analyzed. These themes will be discussed in order of dominance, which means that we will start with the theme that appeared most frequently in all of the participants' responses. Thus, the first theme was the one most frequently mentioned by the respondents.

Familiarity. The first theme that emerged in the arguments raised by the participants is "familiarity". Most participants from the policy, care, and teacher perspectives agree that accommodations should always be allowed in central tests for students with SEN when they are familiar with using them in class. Utilizing these accommodations in the form the students are accustomed to is a logical approach for these students during central tests. Test results are considered valid, and therefore useful for educational development, only if the way in which the national curriculum is assessed in the central tests corresponds to the way in which the students learn the curriculum in the classroom. However, it is noted that effective use in the classroom is a crucial precondition.

> *"This is a normal situation for this child; otherwise, his result will not be representative. You will automatically reduce his chances if you do not allow him to use the tools he is accustomed to in the classroom during the central tests."* (Policy)

Fairness. A second theme that emerged in the arguments raised by participants is "fairness". According to participants representing the legal perspective and to parents, it is fair to *always* allow accommodations in central testing for students with SEN because they are mandated by decree. If students with SEN are not permitted to use accommodations in centralized testing, they cannot demonstrate whether they have mastered the targets, rendering the test results invalid.

There is also consensus among participants from other perspectives that it is fair to consistently allow TTS as an accommodation for SEN if the student is entitled to it. However, it is noted that teachers are not always familiar with information and communication technology (ICT), which sometimes makes them hesitant to use TTS.

> *"We don't have a single student using it. Colleagues are afraid of it because they are not very ICT literate. They are not familiar with it."* (Teacher)

Participants representing the legal perspective state that allowing accommodations in central tests serves as a new lever to encourage teachers to provide students with what they are entitled to. Inclusive practices are compromised if a student with SEN is exempted from the test.

> *"The decree defines refusing accommodations as a form of discrimination. When you, as an employer or educational institution, provide accommodations, the decree does not require justification. It only states that if you refuse them, you must justify why. The fact that accommodations have already been used and applied in education means that they are allowed. If a school grants accommodations, they are already making a decision. If that decision has been made, it is incorrect to withdraw it at the time of the central test. If it is allowed according to the established route, it should not be taken away during tests."* (Legal)

Participants from different perspectives agree that a multilingual student can always use any linguistic support tool to overcome a language barrier when taking central tests, such as a multilingual dictionary or an explanatory dictionary with pictures. It is deemed unfair if these students score lower on items such as number knowledge due to linguistic limitations.

"If linguistic support tools are not allowed in the central tests, it would give a very distorted picture. This also happens in the classroom, so if this is removed in the test, the results are not valid." (Teacher)

"We are very focused on number theory because it is very linguistic. All our students with language problems drop out: multilingual students, students with dysphasia. I notice that if you practice that language very strongly, number theory improves. In numeracy, the language is primarily the stumbling block, not the math. Even with a calculator, they struggle. Some of our secondary school students then use a glossary with examples. They get stuck on the sum, product, quotient, difference. It's about gifted students with this kind of automation disorder."(Policy)

A caveat is made, however, that the use of these linguistic support tools will not (always) resolve the language barrier for multilingual students.

"By making it easier, you're going to simplify content or language structure. Then you're not going to be able to fully measure what a child's abilities are when you have language barriers. …It frustrates me enormously that we can't appreciate children because we always come up against those language barriers. And because we always think we must make it simpler or easier. But in doing so, we set very low expectations for those children. And those children cannot demonstrate what they are capable of. It's not about simplifying but about overcoming language barriers. If your entire central test consists of words you don't understand, I still think that's asking a lot from multilingual children, even if they are normally gifted. They must already have the reading technique and vocabulary grasped before they can understand the text. You already have to understand 98% of the words to comprehend the text correctly." (Care)

Construct validity. A third theme that emerged in the arguments raised by participants is monitoring "construct validity". According to the psychometric, policy, care, and teacher perspectives, accommodations can only be allowed for students to prevent construct-irrelevant variance. The accommodations themselves should not compromise the construct being measured but should only aim to eliminate differences caused by contextual barriers. The interpretation of what is being measured in a particular subskill (e.g., reading comprehension or operations with fractions) varies among participants, influencing their decisions regarding the allowance of accommodations for specific test items. Below are two examples that illustrate this.

In one example, a few participants from the teacher perspective view reducing, making equal, or converting fractions as a form of mental arithmetic. If mental arithmetic is tested, the use of a calculator should not be permitted. Some participants from the care and policy perspectives also feel that a calculator is unnecessary for simple numbers or operations that can be calculated using strategies if more time is allowed. However, most participants from the policy, care and teacher perspectives indicate that a simple calculator with basic functions that respects the order of operations can be used for the central test "operations with fractions", as it alleviates memory load. Students must master the concept of fractions and operations with fractions. If SEN students cannot use a simple calculator for (calculations in) operations, they cannot demonstrate their knowledge and various underlying skills (fairness), according to the care and teacher perspectives.

"When it comes to very simple operations in mental arithmetic, after a while, a little automation can occur. Then a calculator is probably not necessary. But this question concerns operations with fractions, where short-term memory is important and all the steps required must be taken. This is where a calculator is truly needed." (Care)

There is a second example in terms of reading comprehension. According to some teachers, the use of an explanatory dictionary should not be allowed if word strategies are tested (deriving the meaning of a word from context). Some participants from the teacher and policy perspectives also state that if vocabulary or spelling is tested within writing skills, the use of word prediction or a spell-checker should not be permitted, as mastering vocabulary or spelling is part of the construct being measured. However, most participants indicate that SEN students may always use an explanatory dictionary, a spell-checker, and word prediction in the central tests for "reading comprehension" and "writing" to enable them to demonstrate various underlying skills.

Participants offering a psychometric perspective suggest that the construct should be defined as broadly and comprehensively as possible within the context of the core curriculum to ensure that the use of accommodations interferes as little as possible. For example, if reading comprehension focuses on purposeful, functional reading ability—i.e., the purposeful comprehension, interpretation, and evaluation of information within a broad sociocultural context—then the use of linguistic tools is permitted for students who require them. Similarly, if "operations with fractions" focuses on understanding and applying mathematical concepts, a calculator may be used to support computational skills.

Psychometric experts argue that if construct validity is compromised for certain test items, for example, when items are flagged for differential item functioning (DIF), these items should be removed to avoid bias rather than compensating SEN students.

> *"Differential analysis (DIF) can be conducted to see whether the items perform differently when accommodations are used or not. Construct validity can be compromised by allowing accommodations that are not strictly differentially effective and lead to an increase in test performance for all students. In this case, the test items in question should be excluded from the analysis."* (Psychometric)

Comparability. A fourth theme that emerged in some arguments put forward by policy perspective participants is the "comparability" of test scores. They raised concerns that using a formulary may compromise the comparability of results because the content and use of these formularies can vary significantly across schools. They therefore advocate for the use of standard formularies during central tests.

> *"We are also trying to use the same formulary for math in our school. If it works with 10 classes, I think it should be possible to implement it more widely. Many students benefit from a standard formulary, but sometimes you have to make individual adaptations for SEN to truly use it as a tool."* (Policy)

However, the majority of participants across stakeholder groups indicated that a formulary should be individualized and customized. It should be permitted for SEN students, as it serves as a handhold for these students, and they are entitled to use it. Thus, a tension exists among participants as to whether to allow the use of a customized local formula in central tests, depending on whether they prioritize comparability or fairness.

Utility of the accommodation. A final theme that emerged in participants' arguments is "utility". Some teachers question the utility and function of word prediction, and therefore suggest that they should not be permitted in central tests.

> *"The use of word prediction can be rather annoying sometimes, as it often provides incorrect options. It encourages students to write faster. However, with writing skills, you typically want them to check more often. You receive suggestions that are often not helpful."* (Teacher)

How can accommodations be made available to all students as universal tools in digital centralized testing?

Next, participants were asked to what extent accommodations can be made available in the digital testing environment as universal tools for all students. Legal experts indicated that they do not see themselves as experts regarding the use of accommodations as universal tools for all students, and therefore refrained from commenting on the issue.

Several themes emerged from the data set analyzed. Again, the themes are discussed based on how often they were mentioned in the data across all participants. The most frequently mentioned themes are discussed first. With the exception of the theme of "fairness", participants from different perspectives expressed similar ideas on most of the themes, indicating a degree of consensus.

Familiarity. From the perspective of teachers and care professionals, the most important principle is that students participate in centralized testing in a manner consistent with their classroom work. This principle takes precedence over the UDA-principle. Universal tools should not prevent students from receiving individualized accommodations, such as a customized local formulary, which are part of their daily practice.

> *"The very nature of accommodation is that it is individual. There are some things like remediation and differentiation that you can offer in a differentiated learning environment, but the moment you start to compensate, it becomes even more individualized. A one-size-fits-all approach would be quite a step backward for us."* (Care)

Other participant groups also highlighted familiarity as a necessary condition for implementing universal tools. They emphasized that students can only benefit from these tools if they have had sufficient opportunities to practice with them beforehand. Therefore, the range of available tools must be communicated well in advance to allow students to become familiar with them before testing begins.

Fairness. From a psychometric perspective, parents and researchers argued that the most equitable approach is to offer accommodations as digital support tools to all students. Parents believe that universal access ensures that SEN students can use the necessary tools, which they feel is not consistently guaranteed in current school practice. While psychometric experts support universal access, they note that it should apply only to test items for which accommodations are appropriate. When accommodations are permitted, they should be available to all students to avoid creating unintended advantages.

Care coordinators and teachers interpreted fairness differently. They consider the use of accommodations fair if they do not disadvantage students without barriers. They noted that a testing environment in which all digital support tools are available to all students can create tensions when these tools differ from those offered in daily school practice. They argue that unfamiliarity with certain tools can undermine fairness because students who are not accustomed to using a tool may become distracted or confused, especially in the case of younger learners. Therefore, several participants argued that certain tools such as TTS and word prediction, are better suited as individualized accommodations for students who are familiar with them.

> *"If you're not used to working with TTS and word prediction, it can be a huge distraction and then it's counterproductive. Students who aren't accustomed to it will score lower if they are assigned to it. I would really limit this to students who have been assigned this accommodation and are truly familiar with it."* (Care)

Pragmatic Considerations. Participants from both policy and psychometric perspectives suggest that making digital support tools available to all students is practically easier. There is also consensus among all perspectives that providing linguistic support tools to all students is the most practical approach. This is more convenient than allowing for individual provisions, as the latter requires additional registration and administration. Moreover, students who do not need digital support tools will not use them.

Leading Role of the Use of Universal Tools in Central Testing. According to participants offering a policy perspective, the universal use of tools is also trendsetting. It sends a clear message to schools that tools can be utilized in tests and helps to remove the stigma associated with students with SEN. Furthermore, they believe that it is a strategic skill to select and use a tool effectively. In everyday life, various support tools are always available.

> *"The universal principle sends a clear message to schools: test accommodations are possible, even in standardized tests. You are promoting a forward-looking vision here. So, are you going to choose the future or the mainstream? I think this could signify an evolution…. Above all, I believe we should eliminate the stigma by making support tools available to everyone. Experience shows that if students don't need the help, they won't use it."* (policy)

Concerns About Impeding Growth. Some teachers express concern that the universal use of calculators may hinder the development of certain cognitive skills such as computational strategies and logical thinking. Therefore, they are not in favor of the universal use of calculators, preferring to assign them only to SEN students who need them.

> *"What they often must do in grade school is estimate the answer and then see if their answer is close. If they do it mentally, they are likely to be correct much more often. But if they do it with a calculator, they might arrive at an answer that is excessively large and not realize it. They reason much less."* (teacher)

Discussion

This study examined the perspectives of different stakeholders with regard to (1) when accommodations for SEN students should be allowed in digital central tests, and (2) how these can be made available as universal tools for all students. To answer these research questions, focus group discussions were organized to explore various viewpoints.

Accommodations: Tension Between Fairness, Construct Validity, and Comparability

This study revealed significant differences among stakeholder groups, highlighting ongoing tensions regarding the use of accommodations in digital standardized tests in terms of fairness, construct validity, and comparability.

Notably, these differences primarily concerned the rationales underlying the groups' views rather than their ultimate positions on when accommodations should be permitted. Although stakeholders often reached similar conclusions about the appropriateness of accommodations, they did so for different legal, pedagogical, psychometric, or practical reasons.

The stakeholders also invoked different interpretations of fairness. Legal and policy actors framed fairness in terms of mandates—obligations to comply with statutory requirements ensuring accommodations for students with formally recognized special educational needs (SEN). Teachers, care coordinators, and parents, by contrast, associated fairness more strongly with entitlement—a student's right to receive the necessary support to demonstrate their actual competencies. These interpretations differ from the definition of fairness in the testing standards, which emphasize valid score interpretation and reducing construct-irrelevant variance (AERA et al., 2014). However, they illustrate how fairness is understood and implemented across stakeholder groups.

Meanwhile, stakeholders from policy, care, parental and teaching perspectives highlighted the importance of familiarity. They argued that denying students access to the tools they are used to using in the classroom would be both unfair and invalid. According to these participants, familiar accommodations allow SEN students to demonstrate their actual competencies, thereby supporting fairness and construct validity. These participants also stressed the interdependence of fairness and validity. When functional or linguistic barriers prevent a test from accurately measuring a student's ability, the result is invalid and unfair. Accommodations that mitigate these barriers help the test better reflect the intended construct.

Psychometricians and some education professionals cautioned that accommodations could compromise construct validity and result comparability. However, these groups had different interpretations of what constitutes a construct, which led to disagreements about when accommodations distort measurement. From a psychometric standpoint, defining constructs more broadly may reduce the disruptive effect of accommodations.

Concerns were raised about locally developed accommodations such as customized formularies, which vary across schools and may affect standardization. Nevertheless, most stakeholders, particularly those with

policy, care, parental and teaching perspectives, prioritized classroom familiarity over comparability. They contended that accommodations reveal hidden competencies and therefore enhance test validity.

Some teachers questioned the effectiveness of certain tools such as word prediction, suggesting that they might distract students. This hesitancy may reflect limited familiarity with learning disabilities or insufficient training in assistive technologies.

Thus, the broad stakeholder support observed refers to convergence in their overall views, even though the underlying reasoning varied substantially across groups. Further research is recommended to determine if these views are shared in other educational and policy contexts.

From Accommodation to Universal Tool: Concerns about implementation

In response to the second research question, which asked how accommodations could be offered as universal tools in digital assessments, the stakeholders identified different priorities, reflecting broader concerns with regard to implementation.

Familiarity emerged as a key concern, particularly for care coordinators and subject teachers. They emphasized that digital support tools must be integrated into teaching practices because unfamiliar tools can confuse or distract students, especially younger ones. Simply making tools available during tests is not enough; students must be familiar with them in advance.

Fairness and construct validity were also central issues. Parents considered universal access to be fair because it allows SEN students to demonstrate their abilities. However, psychometric experts noted that universal tools are only fair if they are permitted for all students under specific test conditions. Otherwise, there is a risk that some students will gain an unfair advantage. To maintain construct validity, they suggested using differential item functioning (DIF) analysis or exempting relevant items if necessary. However, care coordinators and teachers warned that universal tools should not disadvantage students without learning difficulties.

Policymakers and psychometric stakeholders cited practical utility as a reason to support the use of universal tools because they simplify logistics and reduce administrative burdens. Some policy representatives also emphasized the positive symbolic impact: universal access promotes the strategic use of tools, reduces stigma, and aligns with the goals of inclusive education.

Nevertheless, teachers and care coordinators acknowledged that not all accommodations can be standardized. While a universal toolkit can reduce the need for personalized support, it cannot eliminate it entirely. Some teachers expressed concern that excessive use of support tools (e.g., word prediction) could hinder skill development. They argued for continuing to tailor accommodations on an individual basis.

The concept of fairness revealed internal contradictions among different perspectives. For example, psychometricians feared that SEN students would have an unfair advantage if the support tools were not universal, while teachers were concerned that students unfamiliar with the tools would be at a disadvantage. These tensions underscore the need for further research examining how various stakeholder groups define and implement fairness, as well as how training can address concerns about familiarity and development. It would also be valuable to further investigate how stakeholders' perceptions of the usefulness of specific tools relate to teachers' assumptions that SEN students are unlikely to develop certain educational skills. These assumptions may reflect concerns that accommodations could hinder learning and long-term development. In-depth qualitative research involving teachers, care coordinators, parents, and policymakers could provide insight into how universal accommodations during testing are perceived to influence student growth and learning outcomes.

Furthermore, the role of dictionaries and linguistic support tools remains controversial. Some argue that their universal use undermines construct validity; however, most agree that multilingual students should have

access to linguistic support tools to ensure valid assessment results. However, the appropriateness of such tools is debated. Reviews by Li (2012) and Ríos et al. (2020) show that there is little evidence of performance benefits from the use of bilingual glossaries or dictionaries. De Backer (2020) recommends reducing linguistic complexity in tests, especially in subjects such as mathematics and science, where language should not obscure the content being assessed. This underscores the importance of UDA as a guiding framework for assessment.

Moving Toward Inclusive Assessment: From Medical to Social Perspectives

The structural tensions identified in the earlier analysis, particularly those concerning fairness, construct validity, comparability, and the limits of individualized accommodations, suggest that current assessment practices do not adequately address the diverse needs of learners taking digital standardized tests. These tensions underscore the need for assessment systems that mitigate construct-irrelevant barriers during the design process, instead of relying solely on individualized adjustments. UDA is one such approach, promoting anticipatory and accessible test design to help balance fairness and validity while minimizing the need for retrofitted accommodations (Johnstone, 2003; Thompson, 2002; Smarter Balanced Assessment Consortium, 2014).

The shift toward UDA reflects a broader theoretical change from the medical model of disability, which views disability as an individual deficit requiring accommodation, to the social model, which views disability as resulting from inaccessible environments (Haegele & Hodge, 2016; Liasidou, 2014). In assessment contexts, this perspective emphasizes identifying and removing barriers before testing takes place to ensure that test formats, item types, and digital environments do not unnecessarily disadvantage particular groups of students (Ketterlin-Geller & Johnstone, 2006; Tai et al., 2021).

In this study, stakeholders demonstrated both perspectives. Policymakers, psychometricians, and parents leaned toward the social model, favoring universal tools that promote equity and reduce stigma. In contrast, many teachers preferred individualized accommodations, citing the limited integration of universal tools in the classroom. This result indicates that the individual medical perspective is still prevalent, particularly among teachers. This is not surprising, as teachers are the professionals closest to making accommodation decisions on a regular basis. The individual deficit perspective is embedded in the regulations they must follow, resulting in the medical model dominating assessment practice (Nieminen, 2021; Tai et al., 2021). Accommodations are often mandated by legislation, but students' rights to inclusive and accessible test design are not.

Stakeholders, particularly care coordinators and subject teachers, emphasized familiarity with digital tools as a prerequisite for their effective use in assessments. This aligns with the findings of Lee et al. (2021), who reported minimal use of universal tools among students who were unfamiliar with them. Additionally, the school and the teacher play significant roles in this process. Some teachers in the present study did not find the use of certain accommodations useful or were unfamiliar with ICT, making them more reluctant to use tools such as TTS.

Similarly, a study of educators' perceptions of online accessibility features and their instruction of K-12 English learners (Kim et al., 2022) found that educators and students selectively used certain accessibility features in classroom and testing settings, such as highlighters, line guides, or copy and paste functions. Accessibility features are digital support tools embedded in the test platform that are available to all students, regardless of their learning status. Because they are universally available and do not alter the construct, accessibility features are also commonly referred to as universal tools. Accessibility features that educators perceive as less useful may not be easily contextualized or adapted to classroom instruction. To achieve their intended purpose, accessibility features must be genuinely useful and meaningful to both teachers and learners (Biancarosa & Griffiths, 2012; Ertmer et al., 2012).

Furthermore, this study (Kim et al., 2022) found that barriers to technology integration among teachers or a lack of computers in classrooms limited both students' and teachers' use of accessibility features in instruction and testing. This was also reflected in our study and in a research summary on teachers' perceptions of accommodations (NCEO, 2024). Barriers to technology integration in the classroom may stem from teachers being less familiar with technology or from a lack of systematic support for technology infrastructure and training resources (e.g., Admiral et al., 2017; Bauer & Kenton, 2005; Biancarosa & Griffiths, 2012; Dwyer, 2007; Wozney et al., 2006). This discomfort may influence how teachers perceive and initiate technology integration in the classroom (Bauer & Kenton, 2005).

Previous research on teacher perceptions (Gajria, 1994; Jayanthi et al., 1996; Mathes et al., 2020) has also shown that the perceived usefulness of accommodations is influenced not only by perceived effectiveness but mainly by the feasibility of implementation. The accommodations reported to be the most challenging to implement were those involving technology-based methods.

Limitations and Suggestions for Future Research

This study employed a qualitative approach, enabling a thorough and nuanced examination of stakeholders' perspectives. However, future research should include quantitative methodologies to obtain a more representative and generalizable understanding. Although legal experts participated in this study and offered valuable insights on fairness and legal obligations regarding provisions for students with SEN, they did not contribute to research question 2, which focused on the use of provisions as universal tools for all students. The legal experts noted that this topic was outside their area of expertise. Consequently, the discussion on universal tools lacks legal perspective. This limits the completeness of our findings. Future studies should involve legal experts with background knowledge in accessibility or digital inclusion policy in assessment to make better-informed recommendations on implementing universal tools in tests.

Furthermore, our study included stakeholders with expertise in SEN. However, it is equally important to consider the perspectives of stakeholders without such expertise. Involving a wider range of voices would contribute to a more comprehensive understanding of provisions and the use of universal tools. Future studies could investigate whether the arguments and attitudes of stakeholders differ based on their level of expertise in SEN.

Moreover, as centralized testing is a relatively recent development in Flanders, stakeholders' views may evolve over time as they gain more experience with the system. Longitudinal research could provide valuable insights into how these perspectives shift as familiarity with centralized digital assessments increases. Importantly, future research should also include the voices of students with SEN, who are often best positioned to articulate their own needs, preferences, and experiences (Lazarus et al., 2023).

Further research is needed to understand the long-term effects of accommodations in centralized digital testing environments. A mixed-methods approach combining qualitative and quantitative data could provide a more comprehensive understanding of these effects. In particular, attention should be given to the consequential validity of universal tools, i.e., the intended and unintended consequences of their use on various stakeholders (Messick, 1989; American Educational Research Association, 2014).

In this study, participants representing policy perspectives noted that integrating universal tools into large-scale assessments may inspire more inclusive instructional practices. Therefore, follow-up research should examine how these tools influence educational equity and validity in practice by including the perspectives of students, teachers, parents, support staff, and policymakers. These insights could help determine whether universal tools improve access and enhance the fairness and effectiveness of assessments.

Conclusions and Implications for Policy and Practice

Large-scale digital assessments require decisions that balance fairness, construct validity, comparability, and practical feasibility. This study's findings show that stakeholders approach these issues from different perspectives, leading to structural tensions around the use of accommodations and universal tools. To translate these tensions into actionable guidance for policymakers, a scenario analysis was conducted. This section begins with that analysis and then turns to the broader implications for assessment design, policy, and the implementation of UDA.

Scenario Analysis as a Tool for Evidence-Informed Decision-Making

The findings from this study informed a scenario analysis developed to help policymakers in Flanders determine how to provide digital support tools for centralized testing. The analysis included different scenarios representing a continuum ranging from broad, universal access to digital support tools to highly individualized, accommodation-based models (see Figure 5). When evaluating these scenarios, we took into account the concerns identified in the earlier analysis, particularly those related to fairness, construct validity, comparability, and classroom familiarity. A summary table evaluating the scenarios against these stakeholder concerns is included in Figure 6.

Figure 5. Scenario Analysis to Inform Policy Decisions Regarding Digital Support tools in Standardized Tests

In order to arrive at well-considered policy decisions, different scenarios were developed for the use of content-related digital support tools in standardized tests. These different scenarios can be placed on a continuum from an *open and care-wide use* of digital support tools to the more *individualized provision of tailored accommodations* in standardized tests.

- **Scenario 1: Open and care-wide use of digital support tools**

All digital support tools are made available to all students at all times. Students decide which tools to use and when. During test development, it is explicitly taken into account that all tools are accessible during the assessment, ensuring that construct validity is not compromised.

- **Scenario 2: Limited care-wide use of digital support tools**

All digital support tools are made available to all students; however, access may be restricted for certain test sections or items, depending on the construct being measured. Tool availability is controlled through an on/off function to ensure that tools which compromise construct validity are not enabled. At both the test and item levels, experts determine whether specific tools are appropriate and provide corresponding guidance to schools. If a tool that is normally provided as accommodation for students with SEN is not permitted for a particular item due to construct concerns, those students are exempted from that item.

- **Scenario 3: Individualized open use of digital support tools**

All students have access to the universal tools aligned with the national curriculum. Tool access is again regulated via an on/off function based on the constructs defined in the curriculum standards. Students with SEN receive additional, individually tailored test accommodations.

Variant 3A*: *Individualized open use of digital support tools

Digital support tools not designated as universally available on the centralized testing platform will only be accessible to students with special educational needs (SEN) who are familiar with using them in the classroom. After test administration, Differential Item Functioning (DIF) analysis is used to determine whether specific items function differently for students with and without access to these tools. In this scenario, the use of digital support

tools for certain items is limited to students with SEN, enabling a targeted analysis of potential item bias. Items that display significant DIF should be excluded from final scoring or adjusted for in the analysis.

Variant 3B: Individualized limited use of digital support tools

Students with SEN can use all the test accommodations they are familiar with in education when taking the central test, but not always. Students with SEN are exempt from test where expert judgment has determined in advance that accommodations cannot be allowed due to threats to construct validity.

Figure 6. Different Scenarios for the Use of Digital Support tools in Standardized Testing

Criteria	**S1: Open & care-wide use**	**S2: Limited care-wide use**	**S3A: Individualized open use**	**S3B: Individualized limited use**
Familiarity	Positive	Mixed	Positive	Mixed
Fairness: fair chance to meet standards	Positive	Mixed	Positive	Mixed
Fairness: reduced distraction	Negative	Negative	Positive	Positive
Construct validity: monitored during development	Positive	Not indicated	Not indicated	Not indicated
Construct validity: rules set in advance	Negative	Positive	Positive	Positive
Construct validity: DIF analysis	Not applicable	Not applicable	Positive	Not applicable
Comparability	Positive	Positive	Positive	Positive
Leading role: clear message to schools	Positive	Positive	Positive	Positive
Leading role: strategic skill use	Positive	Positive	Mixed	Mixed
Pragmatic ease	Positive	Negative	Negative	Negative

These scenarios helped education policymakers in Flanders strike a balance between equal opportunities and the validity and reliability of centralized testing. Based on this analysis, Flanders adopted Scenario 3A, which allows for individualized, open digital tool use. Under this model, all students have access to digital support tools related to the curriculum. Access is regulated by an on/off function according to the constructs defined in the curriculum standards. Digital support tools that are not universally available on the testing platform are only accessible to students with SEN who are familiar with using them in the classroom. A post-hoc differential item functioning (DIF) analysis is conducted to monitor item performance and ensure the validity of scores. Items showing significant DIF are excluded from scoring or statistically adjusted.

This scenario analysis illustrates how qualitative insights can inform operational decisions in large-scale assessment systems by translating stakeholder concerns into concrete policy pathways. Thus, it bridges the conceptual tensions identified earlier with the practical demands of designing equitable and valid digital assessments.

Implications for UDA Practice

This study demonstrates that designing centralized digital assessments based on the principles of UDA has great potential for creating inclusive and accessible testing environments. However, its implementation is not without challenges. The findings highlight concerns that should guide the selection of accommodations and universal tools to ensure valid and reliable measurements with minimal barriers. Furthermore, the study reveals that the *social model of disability*, which views disability as a form of human diversity rather than a deficit, has not yet been fully embraced by all stakeholders involved in centralized testing. This suggests the need for a cultural and conceptual shift with regard to how inclusive assessment should be approached.

To strengthen support for UDA among all stakeholders, inclusive testing must be redefined as a *collaborative endeavor* involving students, educators, researchers, school leaders, policymakers, and other key figures (Nieminen, 2022). Inclusive assessment should be understood as a shared responsibility grounded in dialogue and co-construction, not solely as a technical or procedural issue.

This study provides unique, empirically-based insights to inform responsible, evidence-based decision-making regarding the use of accommodations and universal tools in digital centralized assessment. Stakeholders' perceptions of the *face validity* of accommodations revealed broader systemic issues that must be considered when implementing policies regarding the use of both accommodation and universal tools. Thus, the findings meaningfully contribute to the ongoing discourse on establishing inclusive testing environments in large-scale assessments that are fair to all students, while upholding the principles of validity, reliability, and equity.

Practical Recommendations

Based on the results of this study, *practical suggestions* can be formulated for test developers and various stakeholder groups. While these suggestions are not new to the field, they provide a useful summary of recommendations for improving accessibility in testing and assessment consistent with existing literature.

***1. For test developers and psychometricians*: Balance validity and fairness**

Valid measurements and *fair testing* are not mutually exclusive, provided that constructs are carefully defined, tools are applied thoughtfully, and analyses are used to make adjustments as needed. Previous studies (e.g., Cahalan-Laitusis, 2007; Sireci & Scarpati, 2005) have suggested empirical strategies to ensure that scores are comparable when accommodations are used. Our study confirms that stakeholders are aware of the tension between fairness, validity, and standardization. At the same time, they are willing to consider practical, evidence-informed solutions.

- **Define constructs as broadly as possible.**
  Stakeholders emphasized that inclusivity must start at the design stage, with broadly defined constructs to enable the use of content-related support tools without compromising validity (see also Messick, 1989; Sireci & O'Riordan, 2020). This reduces the need to choose between fairness and validity.
- **Allow support tools when they are familiar and do not interfere with the construct.**
  If students are accustomed to using a particular support tool in regular classroom practice, and the tool *does not directly affect* the skill being assessed, then its use should be allowed. This approach supports both fairness and validity.
- **Use DIF analyses to evaluate impact.**
  Psychometric experts emphasized using DIF analyses to evaluate impact. Test items that function differently for students using content-related accommodations (e.g., offering unintended advantages to students without SEN) can be revised or excluded from scoring to avoid bias.
- **Offer targeted exemptions where needed.**

If a content-related support tool *alters* the construct, exempting the student from that specific item may be a valid and fair solution, as outlined in the practical example (see 6.1).

2. *For teachers and support staff*: Strengthen classroom familiarity with digital support tools.

One of the key findings from our focus groups is the importance of being *familiar* with digital support tools. Teachers and care coordinators emphasized that students can only effectively use digital support tools when they are systematically integrated into classroom practice. This finding is consistent with previous research (Lee et al., 2021; Kim et al., 2022) and underscores the need for targeted professional development.

3. *For policymakers and test providers*: Apply clear guidelines and communication regarding permitted support tools.

Clear guidelines and *communication* with regard to permitted support tools are essential for aligning classroom practice with test conditions. Participants, especially school leaders and policymakers, pointed out that unclear guidelines create inconsistencies between schools, and hinder alignment between classroom practice and testing. This finding reinforces the emphasis in the literature on the importance of standardized procedures (AERA et al., 2014; Thurlow et al., 2005; Sireci & O'Riordan, 2020).

4. *For schools and policymakers*: Offer universal tools where possible while maintaining individualized support.

Participants broadly supported a *hybrid model* combining universal tools for all students and individualized accommodations where needed—an approach also supported by the literature (Nieminen, 2022; Thurlow & Kopriva, 2015).

5. *For researchers and policymakers*: Monitor the use and effectiveness of digital support tools.

To fully understand the impact of digital support tools on learning, *longitudinal* and *participatory* research is essential, especially research that includes the perspectives of students with SEN (see Lazarus et al., 2023; De Backer, 2020). Only then can assessment systems become truly inclusive, valid, and equitable.

Received: 12/18/2023. **Accepted:** 11/13/2025. **Published:** 1/2/2026.

Citation: Dierick, S., De Backer, F., & Struyven, K. (2025). Giving voice to stakeholders on fair and valid use of accommodations and universal tools in digital standardized tests. *Practical Assessment, Research, & Evaluation*, 31(1)(1). Available online: https://doi.org/10.7275/pare.3628

Corresponding Author: Sabine Dierick, Hasselt University.
Email: sabine.dierick@uhasselt.be

References

Admiral, W., Louws, M., Lockhorst, D., Paas, T., Buynsters, M., Cviko, A., Janssen, C., De Jonge, M., Nouwens, S., Post, L., Van Der Ven, F., & Kester, L. (2017). Teachers in school-based technology innovations: A typology of their beliefs on teaching and technology. *Computers & Education*, *114*, 57–68. https://doi.org/10.1016/j.compedu.2017.06.013

Ainscow, M., Booth, T., & Dyson, A. (2006). *Improving schools, developing inclusion*. Routledge.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2021). *National protocols for test administration*. https://nap.edu.au/docs/default-source/resources/national-protocols-for-test-administration---naplan-paper-2021.pdf

Baert, H. (2012). *Literature KONDOR project*. Focus groups. Steering Committee Quality Assurance Association KU Leuven.

Bauer, J., & Kenton, J. (2005). Toward technology integration in the schools: Why it isn't happening. *The Journal of Technology and Teacher Education*, *13*(4), 519–546. https://eric.ed.gov/?id=EJ723724

Biancarosa, G., & Griffiths, G. G. (2012). Technology tools to support reading in the digital age. *The Future of Children*, *22*(2), 139–160. https://doi.org/10.1353/foc.2012.0014

Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education*, *25*(3), 141–152. https://doi.org/10.1177/07419325040250030201dolan

Bureau, ICE. (2019). *IEP completion instructions*. https://www.bureau-ice.nl/wp-content/uploads/Afnamehandleiding-IEP-Eindtoets-2019.pdf

Bureau, ICE. (2019). *IEP test regulations-final test*. https://www.Toetsreglement-IEP-Eindtoets-2019.pdf (bureau-ice.nl)

Buzick, H., & Stone, E. (2014). A meta-analysis of research on the read aloud accommodation. *Educational Measurement: Issues and Practice*, *33*(3), 17–30. https://doi.org/10.1111/emip.12040

Cahalan-Laitusis, C. (2007). Differential boost: Examining the interaction of student ability and test accommodation. Educational Testing Service.

De Backer, F. (2020). Multilingual assessment in education: A mixed-methods study of the effectiveness of assessment accommodations and user perspectives on them [Doctoral dissertation, Ghent University]. Ghent University Academic Bibliography. https://biblio.ugent.be/publication/8665597

De Backer, F., Van Avermaet, P., Slembrouck, S., & Van Houtte, M. (2023). Construct-irrelevant variance in testing: A review of dilemmas and empirical strategies. Assessment in Education: Principles, Policy & Practice, 30(2), 135–155. https://doi.org/10.1080/0969594X.2022.2107803

Dolan, R. P., & Hall, T. E. (2001). Universal design for learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22–25.

Dwyer, J. (2007). Computer-based learning in a primary school: Differences between the early and later years of primary schooling. *Asia-Pacific Journal of Teacher Education*, *35*(1), 89–103. https://doi.org/10.1080/13598660601111307

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education, 59*(2), 423–435. https://doi.org/10.1016/j.compedu.2012.02.001

Flemish Ministry of Education and Training. (2017). *M-decree: Broad and inclusive education in Flanders*. Brussels: Flemish Government.

Florian, L. (2014). What counts as evidence of inclusive education? *European Journal of Special Needs Education, 29*(3), 286–294. https://doi.org/10.1080/08856257.2014.933551

Fusch, P., & Ness, L. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2015.2281

Gajria, M., Salend, S. J., & Hemrick, M. A. (1994). Teacher acceptability of testing modifications for mainstreamed students. *Learning Disabilities Research and Practice, 9*(4), 236–243.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing Co.

Hanna, E. I. (2005). *Inclusive design for maximum accessibility: A practical approach to universal design*. Pearson Educational Measurement.

Haegele, J. A., & Hodge, S. (2016). Disability discourse: Overview and critiques of the medical and social models. *Quest, 68*(2), 193–206. https://doi.org/10.1080/00336297.2016.1143849

International Test Commission and Association of Test Publishers. (2022). *Guidelines for technology-based assessment*. https://www.testpublishers.org/white-papers

Joint Council for Qualifications. (2021). *Adaptations for candidates with disabilities and learning difficulties. Access arrangements and reasonable adaptations*. AA_regs_21-22_v11.pdf (jcq.org.uk)

Jayanthi, M., Epstein, M. H., Polloway, E. A., & Bursuck, W. D. (1996). A national survey of general education teachers' perceptions of testing adaptations. *The Journal of Special Education*, *30*(1), 99–115. https://doi.org/10.1177/002246699603000106

Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Technical37.htm

Kettler, R. J. (2012) Testing Accommodations: Theory and research to inform practice. *International Journal of Disability, Development and Education, 59*(1), 53-66. Testing Accommodations: Theory and research to inform practice: International Journal of Disability, Development and Education: Vol 59, No 1

Ketterlin-Geller, L. R., & Johnstone, C. J. (2006). Accommodations and universal design: Supporting access to assessments in inclusive settings. *Teaching Exceptional Children, 38*(3), 65–73. https://doi.org/10.1177/004005990603800310

Kim, A. A., Monroe, M., & Lee, S. (2020). Examining K-12 educators' perception and instruction of online accessibility features. *Computer Assisted Language Learning, 35*(3), 437–468. https://doi.org/10.1080/09588221.2019.1705353

Kim, Y., Kang, S., Yoon, H., & Jun, A. (2022). Teachers' perceptions of online accessibility features for K–12 English learners. TESOL Quarterly, 56(1), 231–258. https://doi.org/10.1002/tesq.3012

Laitusis, C., Buzick, H., Stone, E., Hansen, E., & Hakkinen, M. (2012). *Literature review of testing accommodations and accessibility tools for students with disabilities*. Smarter Balanced Assessment Consortium. Educational Testing Service. Smarter-Balanced-Students-with-Disabilities-Literature-Review.pdf

Lazarus, S. S., Thurlow, M. L., Lail, K. E., & Christensen, L. (2009). A longitudinal analysis of state accommodations policies. *The Journal of Special Education*, *43*(2), 67–80. https://doi.org/10.1177/0022466907313524

Lazarus, S. S., Thurlow, M. L., & Christensen, L. L. (2023). The importance of student voice in assessment accessibility: Emerging evidence from inclusive large-scale assessment research. Journal of Special Education Technology, 38(1), 3–14. https://doi.org/10.1177/01626434221108672

Lee, D., Buzick, H., Sireci, S. G., Lee, M., & Laitusis, C. (2021). Embedded Accommodation and Accessibility Support Usage on a Computer-Based Statewide Achievement Test. *Practical Assessment, Research & Evaluation, 26*(25). http://pareonline.net/getvn.asp?v=26&n=25

Li, H., & Suen, H. (2012). Are test accommodations for English language learners fair? *Language Assessment Quarterly*, *9*(2012), 293–309. https://doi.org/10.1080/15434303.2011.653843

Li, H., & Suen, H. K. (2012). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education, 25*(4), 327-346. https://www.tandfonline.com/doi/abs/10.1080/08957347.2012.714690#.U02voVVdV8E

Liasidou, A. (2014). Critical disability studies and socially just change in higher education. *British Journal of Special Education, 41*(2), 120–135. https://doi.org/10.1111/1467-8578.12063

Mathes, N., Witmer, S., & Volker, M. (2020). Middle school teachers' perceptions of academic and behavioral support testing accommodations. *Journal of Applied School Psychology, 36*(3), 293–323. https://doi.org/10.1080/15377903.2020.1749202

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education and National Council on Measurement in Education.

National Assessment Governing Board. (2014). *NAEP Testing and Reporting on Students with Disabilities and English Language Learners* (nagb.gov). Reporting Test Results for Students with Disabilities and English-Language Learners: Summary of a Workshop | The National Academies Press

National Center for Education Statistics (2019). *NAEP assessment process - Administer the assessment.*

https://nces.ed.gov/nationsreportcard/assessment_process/administer.aspx

National Center for Education Statistics (2021). *NAEP accommodations increase inclusiveness.* NAEP Nations Report Card - Accommodations Table (ed.gov). NAEP Nations Report Card - Accommodations Table

National Center on Education Outcomes (NCEO). (n.d.). *Universal design of assessments.* https://nceo.info/Assessments/universal_design

Nieminen, P. (2021). Framing disability in assessment: A review of Finnish basic education policy. Scandinavian Journal of Educational Research, 65(4), 570–584. https://doi.org/10.1080/00313831.2020.1739137

Nieminen, P. (2022). Universal design in large-scale assessment: Framing inclusive assessment through the lens of validity. Assessment in Education: Principles, Policy & Practice, 29(1), 1–19. https://doi.org/10.1080/0969594X.2021.1993586

Plass, J. L. (Ed.). (2010). Individual differences and cognitive load theory. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 65–87). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.006

Rios, J., Ihlenfeldt, S., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence-based? A multi-study systematic review and meta-analysis. *Educational Measurement: Issues and Practice, 39*(4), 65–75. https://doi.org/10.1111/emip.12337

Rios, J. A., Sireci, S. G., & Talento-Miller, E. (2020). Evaluating the effectiveness and fairness of accommodations for English learners: A review of the literature. *Educational Measurement: Issues and Practice, 39*(2), 36–48. https://doi.org/10.1111/emip.12317

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105–126. https://doi.org/10.1207/s15326977ea1102_2

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. Center for Educational Assessment, University of Massachusetts Amherst.

Sireci, S. G., & O'Riordan, M. (2020). Comparability issues in assessing individuals with disabilities. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability issues in large-scale assessment: Issues and recommendations* (pp. 117–204). National Academy of Education Press.

Sireci, S. G., & O'Riordan, C. (2020). The dilemma of test accommodations and measuring what counts. Educational Psychologist, 55(1), 1–15. https://doi.org/10.1080/00461520.2019.1671806

Smarter Balanced Assessment Consortium. (2014, October 13). *Accessibility and accommodations framework.* https://portal.smarterbalanced.org/library/en/accessibility-and-accommodations-framework.pdf

Solano-Flores, G. (2016). Who Is Given Tests in What Language by Whom, When, and Where? The Need for Probabilistic Views of Language in the Testing of English Language Learners. *Educational researcher, 37*(4), 189-199. https://doi.org/10.3102/0013189X08319569

Tai, J., Ajjawi, R., & Umarova, A. (2021). How do students experience inclusive assessment? A critical review of contemporary literature. *International Journal of Inclusive Education*, 1–18. https://doi.org/10.1080/13603116.2021.2011441

Tai, D. W. S., Chong, S. Y., & Leong, C. H. (2021). Reimagining inclusive assessments: Towards a socially just framework for policy and practice. *Educational Research for Policy and Practice, 20*(2), 123–138. https://doi.org/10.1007/s10671-020-09267-3

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Synthesis44.html

Thurlow, M. L., Lazarus, S. S., Thompson, S. J., & Morse, A. B. (2005). State policies on assessment participation and accommodations for students with disabilities. *The Journal of Special Education, 38*(4), 232–240. https://doi.org/10.1177/00224669050380040401

Thurlow, M. L., & Kopriva, R. J. (2015). Advancing accessibility and accommodations in content assessments for students with disabilities and English learners. *Review Of Research in Education*, *39*(1), 331–369. https://doi.org/10.3102/0091732x14556076

Thurlow, M. L., & Kopriva, R. (2015). Uncovering the relationship between test accommodations and student outcomes. *Educational Measurement: Issues and Practice, 34*(3), 3–12. https://doi.org/10.1111/emip.12058

Thurlow, M. L., Lazarus, S. S., & Christensen, L. L. (2016). Seven principles for inclusive assessment systems. National Center on Educational Outcomes.

Universal Design Project. (2024, October 15). *What is "universal design?"*, Retrieved October 15, 2024, from https://www.universaldesign.org/definition

Vaismoradi, M., Bondas. T., & Turunen, H. (2013). Content Analysis and Thematic Analysis: Implications for Conducting a Qualitative Descriptive Study. *Journal of Nursing & Health Sciences, 15,* 398-405. https://doi.org/10.1111/nhs.12048

UNESCO (1995). *World conference on special needs education: Access and quality Salamanca, Spain, 7–10 June 1994*. Program and meeting document, p. 121.

UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education – All means all.* Paris: UNESCO.

Wozney, L., Venkatesh, V., & Abrami, P. (2006). Implementing computer technologies: Teachers' perceptions and practices. *Journal of Technology and Teacher Education, 14*(1), 173–207.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 131*(1),99-117. https://doi.org/10.1207/s15324818ame1301_5