Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring

Jana Jungjohann[1], Markus Gebhardt[1]

[1] Faculty of Human Science, University of Regensburg, Germany

HOW TO CITE:
Jungjohann, J., & Gebhardt, M. (2023). Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring. *International Journal of Special Education, 38*(1), 131-144.

CORRESPONDING AUTHOR:
Jana Jungjohann;
jana.jungjohann@ur.de

DOI:
https://doi.org/10.52291/ijse.2023.38.12

ABSTRACT:

Inclusive education aims to provide learning progress for all. It requires assessments to identify learning levels and progress and to adapt and evaluate instruction. We extend the classroom-based assessment (CBA) approach and argue that CBA in inclusive schools consists of the four dimensions of Instructional Decision-Making, Educational Assessment, Identification of Students with Special Educational Needs, and Progress Monitoring. For this paper, we developed both a framework of CBA for research interests and individual scales for each dimension (6 to 9 items per scale) in order to identify the four dimensions in which teachers need further training. To create a more manageable instrument for education and training practice, we tested a combined model as a questionnaire. In a survey of 110 teachers and 152 pre-service special education teachers (N = 252), the individual scales had good internal consistency (α = .92, .82, .92, .93). Using Confirmatory Factor Analysis, an initial combined model, consisting of the revised individual scales, showed acceptable fit values (CFI = 0.94, TLI = 0.89, RMSEA = 0.07, SRMR = 0.06) almost everywhere. The further development of the questionnaire and its significance and relevance for inclusive educational practice will be discussed.

Keywords: Classroom-based assessment, Educational assessments, Teacher assessment, Teacher education, Progress monitoring

INTRODUCTION

In inclusive schools, students with and without special educational needs (SEN) should have access to effective education (United Nations, 2006). SEN students need individualized and effective support that maximizes academic and social development (Krämer et al., 2021). Research shows that if children with SEN are only placed, they do not always learn successfully and may be socially excluded (Oh-Young & Filler, 2015). Classroom-based assessment (CBA) aims to shape teaching and learning and is considered one of the crucial components of educational decision-making (Tindal & Marston, 1990; Hill & McNamara, 2011). Its use is intended to improve the learning environment and includes different types of assessments. According to the OECD Report 58 (Looney, 2011), CBA includes both standardized and informal summative and formative assessments. CBA is to be distinguished from a selective or differential assessment. Likewise, no medical model of disability (Gebhardt et al., 2022) is used, but the interaction between individual disposition and the environment is seen as the cause of SEN. In traditional diagnostics, assessment, and support are regarded as separate steps, while in CBA these are closely linked. The focus of CBA is on helping all students meet learning gaps with a close connection to the instructional content and on generating valuable data for decision-makers at the classroom level. Therefore, CBA is seen as part of data-based decision-making (DBDM) to make optimal instructional decisions (Blumenthal et al., 2021; Wilcox et al., 2021). While DBDM refers to the entire decision-making process at all levels from school districts to individual students, CBA remains at the level of selection, implementation, and interpretation of assessment and support within classrooms.

For students at risk or with SEN in inclusive schools, CBA needs to be extended to include identification procedures considering particular learning needs (Silva et al., 2021), which are mainly carried out by special education teachers. The goal of inclusive schooling is for all teachers to know the support aim and interventions for all children, and for instruction to be differentiated in a way that is as individualized as possible. Across countries, two approaches to identifying and supporting students at risk or with SEN exist. In both approaches, the use of CBA is essential but anchored differently. The traditional test-to-diagnosis approach (Fletcher et al., 2019) identifies children with severe school problems through diagnostics to support them individually. This approach is implemented in most countries with a developed special education system. As a consequence, children are then referred to special education or provided with special interventions. In daily classroom practice, regular teachers teach the students and inform special education teachers about the precise academic tasks and students' difficulties (Rasmitadila et al., 2021). The special education teacher initiates an appropriate assessments and then supports the student. This approach can be called a remedial strategy because first a diagnosis is made and then the support is adapted to the diagnosis. The opposite approach is the treat-and-test approach (Fletcher et al., 2019), which focuses on prevention. In the first step, specific literacy difficulties should be monitored and accompanied by formative assessment. Only if this additional support is not effective, further steps in diagnostics and support should be taken. Based on the Response-to-Intervention approach (RTI; Fuchs & Fuchs, 2006), the main idea is that there is no perfect intervention for children with learning difficulties. Therefore, formative assessments are used to find the best support for each student. Both the test-to-diagnosis and the treat-and-test approaches are justified depending on the pedagogical intention, the resources of the school, and the needs of the child. They are part of the diagnostic expertise of regular and special education teachers alike, just as it takes knowledge of effective support methods and expertise in a particular school subject to be a good teacher.

To date, no framework of CBA addresses the requirements of inclusive education. Different facets of diagnostic assessments and steps are summarized under the umbrella term CBA. In inclusive schools, all these facets accumulate due to the different learning needs of the students and the collaboration of regular and special education teachers. Teachers working in inclusive classrooms work hand-in-hand, and their areas of responsibility overlap and complement each other which is why they need broad competencies to use CBA deliberately and purposefully. Existing concepts focus on either content-based teaching or special education (Cumming, 2009; Hill & McNamara, 2011). Therefore, we define a framework of CBA for inclusive education that can be used for the research of teachers' CBA literacy as well as for teacher education and training. We define four dimensions of CBA. (1) **Instructional Decision-Making (IDM)** refers to written language and mathematical literacy development and support strategies for academic skills. (2) The dimension **Educational Assessment (EdA)** describes knowledge about informal and standardized educational assessments for learning and behavior that measures the status quo (i.e., summative assessments).

(3) The third dimension refers to the **Identification of Students with SEN (IdSEN)** which is designed exclusively for special education teachers. (4) In order to document children's learning progress and evaluate support strategies, formative assessments with multiple measurement points are needed which are focused on the dimension **Progress Monitoring (PoM)**. Below, we describe the four dimensions.

Instructional Decision-Making (IDM)

Teaching and support are effective in school when they are adapted to the needs of the child. This is especially true in inclusive schools, where the heterogeneity of achievements is even higher than in regular schools. Additional effective individual support methods are needed for children with learning difficulties and children with disabilities. Difficulties in school may exist in all subjects and for all contents, especially in reading, writing, and mathematics (OECD, 2019). To make effective instructional decisions, teachers need both content knowledge and the pedagogies required to teach the content (Shulman, 2016). Literacy approaches framework relevant steps of learning development and informs teachers what students need to learn in terms of a high level of academic content (McConachie, 2009). These frameworks differ regarding content disciplines and guidance on how to integrate literacy development in their content instruction. For inclusive and special education in particular, the use of evidence-based teaching strategies is recommended for instructional decision-making so that support aims are more likely to be achieved and ineffective learning time is minimized. In sum, each teacher creates their support strategy using different methods, principles, and activities to support their learning group as well as individual students (Sutton & Wheatley, 2003). To improve educational outcomes by choosing adequate support strategies, across subjects, all teachers need to identify students at risk with difficulties emerging in their literacy development. Following literacy approaches, teachers can adapt the instruction to both the students' learning level and individual learning needs. In addition, teachers must evaluate the fit of the support strategy to the students' learning needs to recognize if further adaptions are needed. Evaluation should be based on assessments to improve instruction (Wilcox et al., 2021).

Educational Assessments (EdM)

Data-based decisions need knowledge and competencies in the field of educational assessment in order to be able to carry out and evaluate a measurement. Either informal qualitative or standardized quantitative assessments are used to assess the learning status at a particular point in time (Kibble, 2017). Examples of standardized assessments are state assessments, district benchmarks, or end-of-term screenings which are summative and relative to content standards. To rank students' achievements, standardized assessments can be norm-referenced. They are high-stakes for all concerned and happen far down the learning process. Therefore, students are assessed every few months or once a year to evaluate certain aspects of the learning process, the effectiveness of programs, or student placement. At the classroom level, teachers decide whether to use an educational assessment based on its utility. Van der Vleuten (1996) defined a utility model of educational assessments. The model includes five variables based on which assessment procedures can be accepted: reliability, validity, educational impact, acceptability, and costs. The first three variables relate to scientific quality criteria depending on the purpose of the assessment. Acceptability refers both to the perspective of teachers, which is shaped by traditions and school cultures and to the perspective of the students who are completing an assessment. Therefore, acceptability depends on user-friendliness for both teachers and students. High costs can block the benefits of an assessment completely, even if it is fully appropriate for educational purposes. Many teachers do not have budgets for assessments, so their choices are either limited to the assessments available in the school or resort to free-of-charge assessments. One way of dealing with challenges in terms of non-accessibility or non-fit is for teachers to adapt existing tests. Atjonen (2014) examined the teachers' view of their assessment practice. Results indicate that teachers found it particularly difficult to assess students' outcomes via educational assessments when a) the test fairness was not adequate for all students in the learning group, b) single students needed special education, or c) the achievement range in the class was very heterogeneous. In addition to standardized tests and test adaptions, teachers develop assessments by themselves. Informal tests such as writing samples or multiple-choice questions are often developed to make immediate small decisions (Spinelli, 2008). These self-developed tests run the risk of not measuring student performance reliably and validly. Therefore, it is not recommended to make important decisions based on such results. As with any type of assessment, there is a requirement for teachers to interpret the results concerning the teaching aims.

Identification of students with SEN (IdSEN)

Educational assessments are implemented in most schools in the process of identifying students with SEN (Grünke & Cavendish, 2016). Depending on a child's suspected need for special educational support, the focus of the identification process varies, with all of these processes, however, traditionally using summative assessments such as cognitive, spoken language, reading, or mathematical assessments. Identification is needed to determine the location and extent of support as well as the extent of additional teacher hours. Depending on the local school system, the main responsibility of attribution lies either with a school psychologist, as in the United States (Silva et al., 2021), or with special education teachers, as in Germany (Gebhardt et al., 2021). The persons in charge conduct the assessments with the child, evaluate the results and summarize their observations in a written report (Gebhardt et al., 2013). The individual SEN support measures are also included in this report. Allocations of SEN are negotiated and decided upon jointly by all parties involved (i.e., special education and regular teachers, parents, and child) via the school board based on that written report. Basically, a child receives special educational support after having been examined. Depending on the country, parents are entitled to choose whether their child should attend an inclusive school or a special school (Ebenbeck et al., 2022). Special schools still exist in some countries such as Germany and they are attended only by students with an official SEN diagnosis. These schools were established in the 1960s assuming that they would provide students with SEN with an optimal learning environment, and they are still the majority in Germany. To substantiate parents' choice of school placement, they need information on the advantages and disadvantages of the different types of schools close to home. After deciding on the placement of a child with SEN, it is the task of the special education teacher to design individual education programs for the identified support aims and also check the consistency of the individual child's need for SEN support during instruction in the classroom. If a need no longer exists, the attribution can possibly be lifted. Therefore, special education teachers need more in-depth assessment skills than regular teachers as well as expertise in documenting and reviewing individual education program aims. The role of the special education teacher is to set achievable learning aims even when abilities are low and difficulties are high, and to adjust support and the learning environment with periodic assessments.

Progress Monitoring (PoM)

To measure learning needs, appropriate assessments that measure learning progress reliably and sensitively to enable educational decisions are needed. Formative assessments for progress monitoring provide information at the classroom level to decide about support strategies and to make instructional adjustments during the learning process. Teachers need to know how to translate formative assessment data into support for students. Following Hasbrouck and Tindal (2006), these assessments estimate rates of improvement based on which teachers can identify students who have not achieved the expected progress. Such assessments originate from special education and were heavily researched in the USA then referred to as curriculum-based measurement (Deno, 1985). On the one hand, standardized tests were developed and, on the other hand, guidance for teachers on how to develop their informal formative tests for their specific subject matter was issued (Hosp et al., 2007). Standardized formative assessments meet psychometric quality criteria as well as requirements from school practice, which particularly applies to formative procedures which have more measurement points in order to measure learning progress as accurately as possible. The tests have the shortest possible implementation time of a few minutes and are easy to use for teachers and students (Schurig et al., 2021). Generally, between five to seven measurements are recommended to estimate learning progress reliably. The results are traditionally transferred to a graph to facilitate interpretation by teachers through the visual representation of the data's slope (Jungjohann et al., 2018a). Data interpretation has to be learned to avoid misinterpretations caused by, for example, varying data points or extreme slopes (Klapproth, 2018). Currently, there is still less knowledge of such instruments in school practice than of summative assessments (Blumenthal et al., 2021). Various online platforms and computer-based tests facilitate complex implementations and provide teachers with automatic evaluations. In Germany, for example, there are several paper-pencil tests for primary school (Jungjohann et al., 2018b) and online platforms such as Lernlinien, Levumi, and Quop (Blumenthal et al., 2022). The use of standardized formative assessments is particularly beneficial for students with SEN (Anderson et al., 2020; Stecker et al., 2005). Therefore, the approach is increasingly being applied internationally in different school systems and learning areas by developing and researching appropriate tests (Ardoin, et al., 2013).

CLASSROOM-BASED ASSESSMENT IN TEACHER EDUCATION

A current problem is that the use of CBA is not uniformly regulated, neither in most school systems nor university teacher training programs, resulting in varying degrees of prevalence and inadequate diagnostic expertise (Gebhardt et al., 2015). This is also reflected in standard school practice using assessments. A recent questionnaire study showed that summative screenings are used regularly in over 80% of the schools surveyed, while the use of assessments for progress monitoring was indicated by only about 25% of the schools (Silva et al., 2021). Although the use of data has a positive impact on student achievement, in practice, some teachers still refuse to utilize standard assessment instruments and use their observations instead. These teachers argue that children do poorly on these tests. The comparability and general evaluation of necessary support are therefore seen as negligible by some of the teachers. Changing these attitudes requires rethinking the role of assessment, as Stiggins (1999) called for: "Wise teachers use the classroom assessment process as an instructional intervention to teach the lesson that failure is acceptable at first, but that it cannot continue. Improvement must follow. Success is defined as continual improvement." (p. 196). In the context of language learning, Cumming (2009) formulated three most fundamental issues to develop and improve the use of CBA by teachers. First, professional knowledge and abilities must increase in teacher training and future education. Second, CBA must be connected to relevant school laws and curricula. And third, CBA results must be considered in support planning. Comparable ideas are also disseminated in recent publications (Herppich et al., 2018, Jungjohann et al., 2022, Tra & Linh, 2021).

First of all, teachers need knowledge of how to generate educational data at the level of detail needed to diagnose individual students' needs. Therefore, we see diagnostic competence as a higher-level competency that exists in the four areas mentioned above. We developed a short questionnaire which is based on these four dimensions. All these dimensions are closely related, but independent sections are taught consecutively in teacher training courses in universities or later during teacher training at the schools. The questionnaire is to be used as a whole or in parts for individual scales to identify requirements regarding the diagnostic training of teachers. We evaluated the new questionnaire together with special education teachers to support the implementation of summative and formative assessments in inclusive classrooms.

Research Questions

The present study aims to evaluate the questionnaire and measure the special education, pre-service teachers, and teachers. It is assumed that pre-service teachers have been trained in all dimensions during their studies at university but have less practical experience than in-service teachers. Following the research findings of both Blumenthal et al. (2021) and Silva et al. (2021), all teachers are assumed to know more about summative assessments and less about progress monitoring. The guiding research questions are:

1. Are there differences in the mean values in self-assessment regarding the four instruments IDM, EdA, IdSEN, and PoM between pre-service and in-service teachers?
2. Do the four instruments have acceptable fits in separate confirmatory factor analyses (CFA)?
3. Can the four instruments be combined into a condensed model with acceptable fit values to create a shorter and more manageable instrument for training practice?

METHOD

The survey, dataset, and syntax presented in this study can be found at https://osf.io/exmwq/ (Jungjohann & Gebhardt, 2023).

Instruments

In this study, an online survey was used that included the four individual instruments IDM, EdA, IdSEN, and PoM. In addition, participants were questioned about their age, gender, teacher group (i.e., pre-service teacher, in-service teacher), teacher training (i.e., regular or special education), place of training (i.e., 16 German federal states), prior experience within different school settings (i.e., primary or secondary schools, and inclusive and regular teaching), and number of years in the profession.

The four instruments include six to nine questions asking about key aspects of CBA and have been newly developed (see Table 1). Each scale was presented on a single page within the online survey. Before the first instrument was shown, a note about literacy showed up. Some questions refer to both written language and mathematical literacy. In these cases, teachers should concentrate on the area in which they have more expertise. All questions are asked from the first-person perspective complete with a five-point scale with responses ranging from [1] "not at all true" to [5] "completely true". Thus, participants were asked for their personal agreement.

The first instrument **Instructional Decision-Making** refers to written language and mathematical literacy. It focuses on how to choose, derive, assess, and account for instruction and support for all students with mild learning problems. Therefore, it is suitable for all teachers. For literacy, the extent of the expertise as well as knowledge about teaching and support strategies are surveyed. This instrument contains eight questions.

The second instrument **Educational Assessments** refers to informal and standardized assessments that survey the status quo of school achievements and behavior. Aspects of summative tests selection, use, scoring, and interpretation are considered. Initially, this instrument contained nine questions. This instrument refers to activities conducted by both regular and special education teachers regardless of the school or system.

The third instrument **Identification of Students with SEN** includes only aspects the responsibility of which lies with special education teachers. It asks whether the special education teachers know, administer, and evaluate the necessary steps for SEN identification. Even if the procedure is carried out by a school psychologist, special education teachers need to know and understand all the steps for a comprehensive review of the results. In addition, it is asked whether they can word and check support aims for students with SEN. Since the instrument was initially developed for German teachers, inclusive schools or special school places of support for students with SEN are referenced (for more information about the German school system, see Ebenbeck et al., 2022). This is because parents in Germany are entitled to choose the placement of their children.

The last instrument **Progress Monitoring** refers only to formative assessments. This scale is primarily aimed at special education teachers, but is also suitable for regular education teachers, as formative assessments are also used for students without SEN. Equivalent to the questions of the instrument **AE**, questions are asked about knowledge, selection, use, and interpretation. In addition, the questions ask whether the results of the formative assessments can be transferred to the classroom and the effectiveness of the support in the sense of data-based decision-making.

Participants

In total, 252 special education pre-service teachers and teachers from four German federal states (65% Bavaria, 33.7% North Rhine Westphalia, 0.4% Mecklenburg-Western Pomerania, 0.4% Saxony) took part, with all of them being participants in a teacher training course for progress monitoring. Each of them completed the questionnaire before the start of the training. 152 pre-service teachers (92.0% were female) averaged 26.79 (2.63) years of age and 46% were in their first year of the two-year teacher training. Of the 110 teachers, 90.3 % were female and on average 42.82 (11.16) years old, with a work experience of either 1 - 5 (34%), 6 - 10 (13%), 11 - 15 (13%), or more than 16 years (50%).

Procedures

Mean score comparison

For every single item, the mean score (M), and standard deviation (SD) separately for both groups of participants (i.e., pre-service teachers and teachers) as well as Cohen's D for quantifying the effect size of the comparison were examined. The effect size of 0.2 can be interpreted as small, 0.5 as medium, and 0.8 as large (Cohen, 1988). For each scale, a composite mean score comprising all items, Cronbach's alpha to describe the internal consistency, and the corrected item-total correlations (i.e., correlation of that item with the scale total if that item is not included in the scale; r_{it}) were calculated.

Group comparisons were made by *MANOVA* between the four composite mean scores of the two groups pre-service teachers and teachers. Partial eta squares of .01, .06, and .14 correspond to small, medium, and large effects (Cohen, 1988).

Factor analyses

Separate CFAs were calculated with laavan (Rosseel, 2012) in GNU R (R Core Team, 2020) using robust maximum likelihood estimation (MLR). No residual correlations were allowed by any model. Persons were excluded listwise due to missing values (i.e., 10 pre-service teachers, 13 teachers). Following Hu and Bentler (1998), we use the following cutoff values to assess the appropriate model fit. An acceptable model fit shows values of RSEMA <0.08, CFI >0.90, TLI >0.90, and SRMR <0.08, while a model fit is considered as good with values of RSEMA <0.05, CFI >0.95, TLI >0.95 and SRMR <0.05. In line with Kenny et al. (2015), regarding the RMSEA, we expect them to be potentially misleading due to the low degrees of freedom. We follow the recommendations of Kenny et al. (2015) and will not reject model fit based solely on the RMSEA. Thus, we decide on the acceptance of the model based on CFI, TLI, and SRMR.

In addition, two combined models were developed which incorporated the modified structures of every single scale. The first combined model M1 includes two

latent factors. The one latent factor unifies the instruments **IDM** and **EdA** as they describe the activities of all teachers regardless of their profession. Their content can be considered basic knowledge of all teachers. The other latent factor includes the other two scales **IdSEN** and **PoM** because the assessment competencies included are primarily carried out and guided by special education teachers. In the second combined model M2, each scale was represented as a single latent factor. All scales were allowed to co-vary with each other. This allowed us to use the co-variances to examine the hypothesized relationships between each scale.

RESULTS

Mean score comparisons

Reliability coefficients ranged between α = 0.82 - 0.96, meaning the overall scale and the single subscales were highly reliable in terms of internal consistency. In the three subscales **IDM, IdSEN,** and **PoM**, all items show sufficiently good correlations with the overall scale (**IDM**: r_{it} = 0.61 - 0.82, **IdSEN**: r_{it} = 0.71 - 0.85, **PoM**: r_{it} = 0.69 - 0.88). In the scale **EdA**, item **EdA04** does not correlate sufficiently with the scale overall (r_{it} = 0.29). Following Field et al. (2012), this item will be dropped. With the item **EdA04** removed, the remaining items of the scale **EdA** also achieve sufficient correlations with the entire scale (**EdA**: r_{it} = 0.43 - 0.77).

The descriptive results from Table 1 show that almost all mean values were in the upper half of the scale. In particular, higher ceiling effects can be observed among the group of teachers than that of the pre-service teachers. Teachers rate their competencies in all subscales at the upper end of the scale and higher. Using the Pillai trace, a one-factor MANOVA revealed that significant differences were present in the four composite mean scores across pre-service teachers compared to teachers with a large effect (V = .62, $F(1, 250) = 101.81$, $p < .001$, $\eta^2 = 0.29$). While the effects in all scales can be classified as large (η^2 = 0.19 - 0.62), the largest effect relates to scale **IdSEN** ($\eta^2 = 0.62$).

Table 1 **Item characteristics separated by the four scales**

No.	Item description	Pre-T[1]	In-T[2]	Effect size	All T[3]
		M (SD)	M (SD)	*d*[4]	r_{it}
I Instructional Decision-Making (α = .92, *N* = 252)					
01	I know how to identify students with learning problems.	3.91 (0.70)	4.55 (0.56)	0.99	0.62
02	I am familiar with literacy approaches (i.e., reading, writing or mathematic).	3.67 (0.78)	4.27 (0.81)	0.76	0.61
03	I am familiar with methods of how to support literacy development in reading, writing or mathematics.	3.58 (0.68)	4.31 (0.72)	1.06	0.76
04	I know how to choose a suitable support strategy for a child.	3.39 (0.70)	4.18 (0.81)	1.06	0.80
05	I know how to adapt a support strategy for literacy development.	3.17 (0.78)	4.05 (0.91)	1.06	0.77
06	I know how to evaluate the level of fit of a specific support strategy for a child.	3.16 (0.76)	4.11 (0.86)	1.19	0.82
07	I am familiar with developing individualized material to be used for support independently.	3.30 (0.85)	4.25 (0.86)	1.12	0.70
08	I know how to assess the reading, writing or mathematical literacy of a child when writing a school report.	3.37 (0.79)	4.34 (0.70)	1.30	0.71
	Composite	3.44 (0.75)	4.26 (0.78)	1.46	
II Educational Assessment (α = .82, *N* = 252)					
01	I know several informal and standardized tests for measuring the status quo (summative assessments).	3.55 (0.84)	4.14 (0.92)	0.68	0.62

02	I am familiar with norm-reference values and their relevance in educational assessment.	3.57 (0.89)	4.18 (0.77)	0.72	0.60
03	I choose an educational assessment based on its quality criteria.	2.96 (0.98)	3.31 (0.98)	0.36	0.43
04	I choose an educational assessment test based on its user-friendliness.	3.57 (0.83)	3.78 (0.79)	0.27	0.29
05	I know how to conduct informal and standardized tests for measuring the status quo of students and how to evaluate them.	3.61 (0.78)	4.42 (0.73)	1.07	0.70
06	I know how to interpret results from standardized tests for measuring the status quo and how to assess educational aims.	3.46 (0.83)	4.22 (0.73)	1.02	0.77
07	I know how to design informal tests for measuring the status quo independently (e.g., reading comprehension or number range extension tests).	2.53 (0.67)	3.58 (1.03)	1.03	0.52
08	In everyday life at school, I know how to quickly and easily get a comprehensive overview of the achievement level of my class even without tests.	3.30 (0.56)	3.83 (0.88)	0.62	0.43
09	If, in specific cases, there is no suitable standardized educational assessment available, I know how to adapt and use a standardized assessment instrument for concrete purposes (e.g., choosing subtasks).	2.93 (0.70)	3.56 (1.00)	0.67	0.59
	Composite	3.28 (0.87)	3.89 (0.87)	1.16	
III Identification of Special Educational Needs (α = .92, *N* = 252)					
01	I have ample experience in writing special education reports.	1.82 (0.80)	4.13 (1.09)	2.51	0.78
02	I am familiar with the steps necessary to be taken to assess a child's need for special educational support.	2.86 (0.85)	4.62 (0.65)	2.27	0.85
03	I know how to conduct assessments to assess a child's need for special educational support.	2.84 (0.88)	4.70 (0.63)	2.37	0.85
04	I know how to phrase the aims of special educational support.	3.57 (0.86)	4.74 (0.54)	1.56	0.79
05	I know different placements (also inclusive ones) and am familiar with providing advice on their advantages and disadvantages.	3.38 (0.93)	4.53 (0.82)	1.30	0.71
06	I know how to check whether the aims of a special support strategy have been achieved.	3.43 (0.80)	4.48 (0.67)	1.40	0.77
	Composite	2.98 (0.85)	4.53 (0.73)	2.60	
IV Progress Monitoring (α = .93, *N* = 252)					
01	I am familiar with different instruments for progress monitoring (formative assessments).	2.63 (0.84)	3.37 (1.03)	0.80	0.79
02	I am familiar with the theoretical background and concepts of progress monitoring approaches.	2.60 (0.88)	3.28 (1.13)	0.69	0.76
03	I know how to conduct and evaluate a progress monitoring instrument.	2.74 (0.93)	3.56 (1.09)	0.83	0.88
04	I know how to interpret student progress monitoring data presented graphically.	2.73 (1.00)	3.58 (1.06)	0.84	0.82
05	I know how to adapt instruction based on student progress monitoring data.	2.89 (0.99)	3.71 (1.04)	0.82	0.86

06	I am familiar with designing an informal instrument for progress monitoring.	2.28 (0.91)	3.16 (1.10)	0.89	0.74
07	I know how to evaluate the effectiveness of a support strategy.	3.05 (0.95)	3.84 (0.83)	0.87	0.69
	Composite	2.71 (0.93)	3.50 (1.04)	1.00	

Note. *[1] Pre-T. = Pre-service teachers (n = 152). [2] In-T. = In-service teachers (n = 100). [3]All T. = All special education teachers (N = 252). [4] Cohen's d.*

Confirmatory factor analysis of the single scales

For CFA, two items (**EdA04**, **EdA08**) were excluded both of which simultaneously show low correlation values for the dimensions as well as a low effect size in the mean comparisons of the participated groups. In Table 2, the model fits for the individual instruments and for the combined models M1 and M2 are presented. Overall fits ranged from acceptable to good. Modifications were needed in the **EdA** and **IdSEN** scales which are described below. In addition, the RMSEA exceeded the recommended threshold according to Hu and Bentler (1999). Due to the small sample size and the resulting low degrees of freedom of the models, the RSMEA is not meaningful (Kenny et al., 2015). For completeness, the RMSEA is also given in Table 2.

Instructional Decision-Making

The initial **IDM** instrument produced fits ranging from good to acceptable. Therefore, no modifications were made.

Educational Assessment

The instrument **EdA** had an acceptable value for SRMR, but CFI and TLI values were too low. We allowed for a pairwise measurement error correlation between **EdA07** and **EdA09**. Both items refer to the fit of tests for their purpose. **EdA07** asks about competence assessment with regard to test redesign and **EdA09** asks about the ability to adapt a test for one's purpose. Therefore, it is likely to share some variance. The indices of modification reinforce this assumption. This modification produced a model with an acceptable to a good fit.

Identification of students with SEN

The initial **IdSEN** scale had a good value for SRM, an acceptable value for CFI, but a value for TLI that is too low. Similar to the **EdA** scale, modifications were necessary. Two items explicitly relate to the support aims of special education. **IdSEN04** asks about phrasing and **IdSEN06** about reviewing an aim. This relationship was also reinforced by

Table 2. **Model fits of separate scales and combined models**

Model	**RMSEA (90% CI)**	**CFI**	**TLI**	**SRMR**
IDM	0.114 (0.092 – 0.136)	0.932	0.905	0.050
EdA	0.131 (0.107 – 0.156)	0.882	0.824	0.076
EdA, modified	0.096 (0.070 – 0.123)	0.941	0.905	0.055
IdSEN	0.170 (0.139 – 0.203)	0.918	0.863	0.047
IdSEN, modified	0.114 (0.081 – 0.150)	0.967	0.938	0.040
PoM	0.131 (0.104 – 0.159)	0.946	0.919	0.037
Combined model M1	0.110 (0.105 – 0.116)	0.770	0.750	0.086
Combined Model M2	0.072 (0.066 – 0.078)	0.940	0.893	0.062

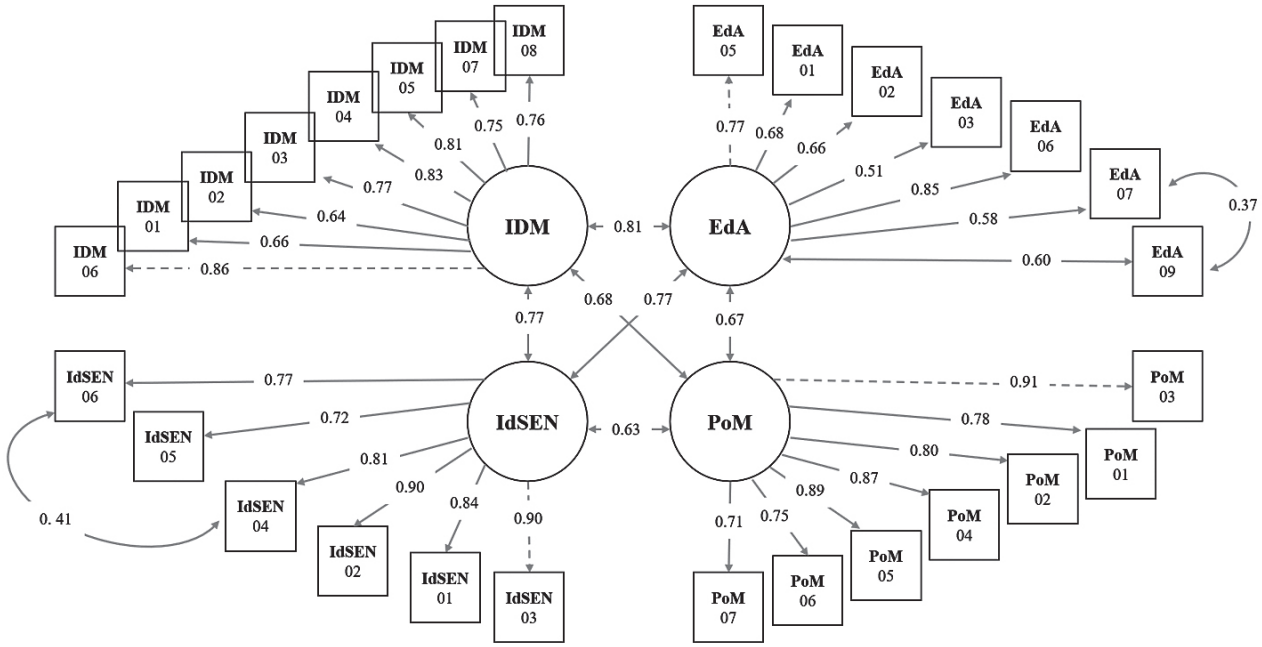

Fig. 1. Combined Model M2 with Path Loading

the modification indices. Therefore, we also allowed these items to co-vary. This modification produced a good model.

Progress Monitoring

The **PoM** model produced acceptable to good fits. No modifications were required.

Combined Models

In the first combined model **M1** with two latent factors, no value had an acceptable fit. Therefore, this model is not sustainable. In the second combined model **M2**, only the TLI value was just below the threshold. In total, we conclude that the overall fit of model **M2** is acceptable. In addition, changes in the combined models would lead to possibly divergent structures concerning the latent variables compared to the individual models. Therefore, we decided to keep the implicit structure of the combined model **M2** and not make any further modifications. In Figure 1, the four-dimension model **M2** is shown. The standardized factor loadings range from 0.51 to 0.91.

DISCUSSION

Students with and without SEN require appropriate support (Krämer et al., 2021). Such support is only possible with trained teachers, sufficient time, and the implementation of CBA in the school. So far, CBA is more or less taken for granted in everyday educational practice, but its actual use depends greatly on the individual teacher. In our study, we developed and successfully evaluated a new questionnaire for assessing the requirements in teacher training related to CBA. The instrument showed good reliable values in the self-assessment of the special education teachers. Therefore, it can be used in practice with all four dimensions or only with one single dimension if necessary. The theoretical assumptions of the four dimensions of CBA could be confirmed empirically. This is probably due to the fact that the framework of the four instruments is very much oriented toward the curriculum of the universities and the daily work of special education teachers. The comparison of the responses of the special education pre-service and in-service teachers showed that each instrument could measure expected lower mean values by the more inexperienced teachers with large effects (η^2 = 0.19 - 0.62). The largest effect relates to the scale **IdSEN** (η^2 = 0.62) which can be explained by the different levels of experience of the participants. Activities considered in that scale are explicit content of the second phase of the teachers› training in which the pre-service teachers were at the time of the survey. Thus, they had not yet finalized this training content. However, the lowest competencies attributed by all teachers with the smallest effect was observed in the scale **PoM** (η^2 = 0.19), indicating low experiences of both groups in this area. This was to be expected, since this is a new area in Germany (Blumenthal et al., 2021) and will only be taught on a larger scale at universities and implemented in practice in the next few years (Jungjohann et al., 2018b).

A high level of understanding of the items could be achieved in almost all cases. Three items out of a total of 30 were misleading one of which was the **EdA** scale. The first conspicuous item refers to the importance of the classic quality criteria of a scientific assessment procedure (**EdA03**). Following the model of the utility of educational assessment by van der Vleuten (1996), teachers should consider quality criteria for choosing an assessment. The item showed both low correlation values and effect size in the mean comparisons of the participating groups. Since the quality criteria are indispensable for the selection of a test procedure, it was retained despite the low fit values. Although quality criteria provide teachers with an easily accessible criterion of objectivity, reliability, and validity (van der Vleuten, 1996), they do not ascribe importance to them when selecting a test procedure in our study. Since it is the task of university training to convey the importance of quality criteria to teachers, a need to catch up can be deduced here. Only when teachers have understood and internalized the range of statistical values can they be used as a basis for selection.

Potentially due to difficulties in understanding, two other items did not fit well. They were removed from the **EdA** scale. The item **EdA04** that had first been removed related to the usability of an assessment. The wording may be too general since teachers have probably neither a uniform understanding of the term user friendliness nor an awareness of its relevance in the application. From a scientific perspective, usability has become a recognized secondary quality criterion, especially for formative tests (Schurig et al., 2021). With this wording, it was not apparent to the teachers whether user-friendly referred to themselves or the handling of the tasks for the students, as van der Vleuten (1996) differentiated it. The wording of the second item **EdA08** that was removed was probably insufficiently stringent. Thus, our interpretation is that the use of several adjectives leads to different interpretations. One consequence could be that both very competent and rather incompetent teachers may rate their abilities as very high.

The results of the study are limited in several places. As in many questionnaire studies, we had to use an ad-hoc sample which is not representative. All participants first registered for training on progress monitoring and were then asked to voluntarily participate in the questionnaire before the training began. Therefore, it may well be that they have a particular interest in CBA. This might bias the representativeness of the results. A second limitation may be due to the theoretical derivation of the CBA framework. The framework was created in such a way that it can also be applied in Germany. Since special education in the German school system is different from that in other school systems (Grünke & Cavendish, 2016), the transferability of the results may be limited. Depending on how the identification of SEN processes and the responsibilities in inclusive education are regulated, the instruments would have to be adapted to local conditions. Methodologically, the sample size and degrees of freedom limit the results. The combination of small sample size and low degrees of freedom often results in CFA that do not reach the cutoff values. This is especially true for the RSMEA, which was also observed in our study. Kenny et al. (2015) have suggested that under these conditions greater importance should be attached to the other fit values. We followed this recommendation. Nevertheless, a larger sample should be used in the test to see whether the model fit improves.

Education training in the field of CBA is steadily on the increase as the demand for evidence-based decision-making grows. To better tailor these offers to the target group, questionnaires such as the one presented here, are an important tool for practice and research. Further development of the questionnaire is therefore necessary. On the one hand, the misfitting items must be redesigned to be closer to the intended latent construct. On the other hand, the questionnaire must be evaluated with regular teachers who work in inclusive classrooms together with special education teachers. Overall, the questionnaire shows that there is a need for good and regular training of practitioners. Thus, there should also be in-service and further training for practicing teachers in schools.

ACKNOWLEDGEMENT

None

DECLARATION OF INTEREST STATEMENT

No potential conflict of interest was reported by the author.

FUNDING

None

ETHICS APPROVAL

The research was conducted in accordance with APA ethical procedures and guidelines. Participants provided informed written consent and could terminate participation at any point without any disadvantage to themselves and data were treated confidentially. Because no experimental manipulation, invasive procedure, or highly sensitive data were collected, a full ethics application from the University Ethics Committee was not required.

REFERENCES

Anderson, S., Jungjohann, J., & Gebhardt, M. (2020). Effects of using curriculum-based measurement (CBM) for progress monitoring in reading and an additive reading instruction in second classes. *Zeitschrift Für Grundschulforschung*, 13(1), 151–166. https://doi.org/10.1007/s42278-019-00072-5

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology,* 51(1), 1–18. https://doi.org/10.1016/j.jsp.2012.09.004

Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal,* 25(2), 238–259. https://doi.org/10.1080/09585176.2013.874952

Blumenthal, S., Blumenthal, Y., Lembke, E. S., Powell, S. R., Schultze-Petzold, P., & Thomas, E. R. (2021). Educator perspectives on data-based decision making in Germany and the United States. *Journal of Learning Disabilities*, 54(4), 284–299. https://doi.org/10.1177/0022219420986120

Blumenthal, S., Gebhardt, M., Förster, N., & Souvignier, E. (2022). Internetplattformen zur Diagnostik von Lernverläufen von Schülerinnen und Schülern in Deutschland. Ein Vergleich der Plattformen Lernlinie, Levumi und quop. *Zeitschrift für Heilpädagogik,* 73(4), 153–167.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Cumming, A. (2009). What needs to be developed to facilitate classroom-based assessment? *TESOL Quarterly*, 43(3), 515–519. https://doi.org/10.1002/j.1545-7249.2009.tb00247.x

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52(3), 219–232. https://doi.org/10.1177/001440298505200303

Ebenbeck, N., Rieser, J., Jungjohann, J., & Gebhardt, M. (2022). How the existence of special schools affects the placement of students with special needs in inclusive primary schools. *Journal of Research in Special Educational Needs,* 22(3), 274–287. https://doi.org/10.1111/1471-3802.12565

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles: Sage.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). New York: Guilford Publications.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93–99. https://doi.org/10.1598/RRQ.41.1.4

Gebhardt, M., Krammer, M., Schwab, S., Rossmann, P., & Gasteiger-Klicpera, B. (2013). What is Behind the Diagnosis of Learning Disability in Austrian Schools? An Empirical Evaluation of the Results of the Diagnostic Process. *International Journal of Special Education,* 28(3), 147–153.

Gebhardt, M., Schwab, S., Krammer, M., & Gegenfurtner, A. (2015). General and special education teachers' perceptions of teamwork in inclusive classrooms at elementary and secondary schools. *Journal for Educational Research Online*, 7(2), 129–146. https://doi.org/10.25656/01:11493

Gebhardt, M., Jungjohann, J., & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht: Testkonstruktionen, Instrumente, Praxis [Curriculum-based Measurement in support-oriented teaching: test constructs, instruments, practice].* Munich: Reinhardt Verlag.

Gebhardt, M., Schurig, M., Suggate, S., Scheer, D., & Capovilla, D. (2022). Social, Systemic, Individual-Medical or Cultural? Questionnaire on the Concepts of Disability Among Teacher Education Students. *Frontiers in Education, 6*, Article 701987. https://doi.org/10.3389/feduc.2021.701987

Grünke, M. & Cavendish, W. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal*, 14(1), 1–8.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59(7), 636–644. https://doi.org/10.1598/RT.59.7.3

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education,* 76, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. https://doi.org/10.1177/0265532211428317

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABC's of CBM: A practical guide to curriculum-based measurement (1st ed.). The Guilford Practical Intervention in the Schools.* New York: The Guilford Press.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Jungjohann, J., & Gebhardt, M. (2023). *Dimensions of Classroom-Based Assessments in Inclusive Education: Data & Syntax.* https://doi.org/10.17605/OSF.IO/EXMWQ (access: 2023/04/12).

Jungjohann, J., Gebhardt, M., & Scheer, D. (2022). Understanding and improving teachers' graph literacy for data-based decision-making via video intervention. *Frontiers in Education,* 7, 919152. https://doi.org/10.3389/feduc.2022.919152

Jungjohann, J., Diehl, K., Mühling, A., & Gebhardt, M. (2018a). Graphen der Lernverlaufsdiagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi [Interpret and use curriculum-based measurement graphs - Reading interventions with online platform for progress monitoring called Levumi]. *Forschung Sprache,* 6(2), 84–91. https://doi.org/10.17877/DE290R-19806

Jungjohann, J., Gegenfurtner, A., & Gebhardt, M. (2018b). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit [Systematic review of learning progress monitoring in early reading fluency]. *Empirische Sonderpädagogik,* 10(1), 100–118. https://doi.org/10.25656/01:15963

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. https://doi.org/10.1177/0049124114543236

Kibble, J. D. (2017). Best practices in summative assessment. *Advances in physiology education*, 41(1), 110–119. https://doi.org/10.1152/advan.00116.2016

Klapproth, F. (2018). Biased predictions of students' future achievement: An experimental study on pre-service teachers' interpretation of curriculum-based measurement graphs. *Studies in Educational Evaluation,* 59, 67–75. https://doi.org/10.1016/j.stueduc.2018.03.004

Krämer, S., Möller, J., & Zimmermann, F. (2021). Inclusive education of students with general learning difficulties: A meta-analysis. *Review of Educational Research,* 91(3), 432–478. https://doi.org/10.3102/0034654321998072

Looney, J. W. (2011). Integrating formative and summative assessment: Progress toward a seamless system? *OECD Education Working Papers*, 58. https://doi.org/10.1787/5kghx3kbl734-en

McConachie, S. M. (2009). *Content matters: A disciplinary literacy approach to improving student learning. The Jossey-Bass education series.* Hoboken: John Wiley & Sons.

OECD. (2019). *PISA 2018 Results (Volume I): What students know and can do.* OECD Publishing and Centre for Educational Research and Innovation. https://doi.org/10.1787/5f07c754-en

Oh-Young, C., & Filler, J. (2015). A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities. *Research in Developmental Disabilities*, 47, 80–92. https://doi.org/10.1016/j.ridd.2015.08.014

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from: https://www.R-project.org/ (access: 2023/04/12).

Rasmitadila, R., Tambunan, A. R. S., Achmadtullah, R., Nuraeni, Y., Samsudin, A., & Nurtanto, M. (2021). Teachers' instructional interaction in an inclusive classroom: Interaction between general teacher and special assistant teacher. *International Journal of Special Education*, 35(1). https://doi.org/10.52291/ijse.2020.35.2

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). https://doi.org/10.18637/jss.v048.i02

Schurig, M., Jungjohann, J., & Gebhardt, M. (2021). Minimization of a short computer-based test in reading. *Frontiers in Education*, 6, 1-12. https://doi.org/10.3389/feduc.2021.684595

Shulman, L. S. (2016). *What teachers should know and be able to do*. Arlington: National Board for Professional Teaching Standards.

Silva, M. R., Collier-Meek, M. A., Codding, R. S., Kleinert, W. L., & Feinberg, A. (2021). Data Collection and Analysis in Response-to-Intervention: a Survey of School Psychologists. *Contemporary School Psychology,* 25(4), 554–571. https://doi.org/10.1007/s40688-020-00280-2

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools,* 42(8), 795–819. https://doi.org/10.1002/pits.2011

Spinelli, C. G. (2008). Introduction: The benefits, uses, and practical application of informal assessment procedures. *Reading & Writing Quarterly*, 24(1), 1–6. https://doi.org/10.1080/10573560701753005

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191–198.

Sutton, R. E. & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review,* 15(4), 327–358. https://doi.org/10.1023/A:1026131715856

Tindal, G. A., & Marston, D. B. (1990). *Classroom-based assessment: Evaluating instructional outcomes*. Columbus: Merrill Publishing Co.

Tra, D. H., & Linh, N. T. D. (2021). A competence model to assess and develop designing competence assessment tool. *International Journal of Learning, Teaching and Educational Research,* 20(2), 81–103. https://doi.org/10.26803/ijlter.20.2.5

United Nations, (2006). *Convention on the rights of persons with disabilities*. Retrieved from: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html (access: 2023/04/12).

Van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications*. Advances in Health Sciences Education,* 1(1), 41–67. https://doi.org/10.1007/BF00596229

Wilcox, G., Fernandez Conde, C., & Kowbel, A. (2021). Using evidence-based practice and data-based decision making in inclusive education. *Education Sciences*, 11(3), 129. https://doi.org/10.3390/educsci11030129