Revue phénEPS / PHEnex Journal

Vol 16 no 1

Development of a Differentiated Instruction Practices Questionnaire in Physical Education: A First Comprehensive Phase in a French-Canadian Context

Claudia Verret
Université du Québec à Montréal
Montréal, Québec
CANADA

Stéphanie Girard
Université du Québec à Trois-Rivières
Trois-Rivières, Québec
CANADA

Maxime Mastagli
Université de Lorraine
Metz
FRANCE

Line Massé
Université du Québec à Trois-Rivières
Trois-Rivières, Québec
CANADA

Geneviève Bergeron
Université du Québec à Trois-Rivières
Trois-Rivières, Québec
CANADA

Author Biographies

Claudia Verret (PhD), is a Professor in the Department of Physical Activity Sciences at the Université du Québec à Montréal.

Stéphanie Girard (PhD), is a Professor in the Department of Human Kinetics at the Université du Québec à Trois-Rivières.

Maxime Mastagli (PhD), is an Associate Professor in the Laboratory of Psychological and Neuroscience Behavioral Dynamics at the Université de Lorraine

Line Massé (PhD), is a Professor in the Department of Psychoeducation and Social Work at the Université du Québec à Trois-Rivières.

Geneviève Bergeron (PhD), is a Professor at the Department of Educational Sciences at the Université du Québec à Trois-Rivières.

Correspondence concerning this article should be addressed to Claudia Verret, Ph.D. Department of Physical Activity Science, Université du Québec à Montréal, Montréal, Canada.
E-mail: verret.claudia@uqam.ca

Abstract

This study sought to develop and provide an initial psychometric validation of a questionnaire measuring differentiated instruction (DI) practices in physical education (PE) following Boateng et al.'s (2018) validation framework. With 240 PE teachers (130 women; *M*_experience = 15.5 years) from primary (72%) and secondary schools (28%), principal component and confirmatory factor analyses advocate a four-factor structure comprising 16 items across frequency and competency scales: (1) student heterogeneity, (2) autonomy and motivation practices, (3) handling heterogeneity practices, and (4) student progress monitoring. Descriptive results showed moderate to high self-reported scores, with content adjustments most frequently implemented and cultural responsiveness least frequent. Preliminary evidence of concurrent validity with attitudes towards inclusion and measurement invariance across gender was observed. This initial psychometric validation in a French-Canadian context addresses a measurement gap in PE literature. These preliminary findings indicate promise for the instrument's use while highlighting the need for continued validation across diverse contexts, longitudinal designs, and predictive validity studies.

Keywords**:** inclusion; teaching practices; heterogeneity; self-report questionnaire; competency; scale development

Résumé

Cette étude vise à développer et à procéder à une validation psychométrique initiale d'un questionnaire mesurant les pratiques de différenciation pédagogique (DP) en éducation physique et à la santé (EPS), selon le cadre de validation de Boateng et al. (2018). L'échantillon se compose de 240 personnes enseignantes d'EPS (130 femmes; *M*_expérience = 15,5 ans) du primaire (72%) et du secondaire (28%). Des analyses en composantes principales et factorielles confirmatoires soutiennent une structure à quatre facteurs comprenant 16 items répartis sur des échelles de fréquence et de compétence : (1) l'hétérogénéité des élèves, (2) les pratiques d'autonomie et de motivation, (3) les pratiques de gestion de l'hétérogénéité et (4) le suivi des progrès des élèves. Les résultats descriptifs révèlent des scores auto rapportés modérés à élevés. Les ajustements de contenu sont les pratiques les plus fréquemment mises en œuvre et la sensibilité culturelle, la moins fréquente. Des données préliminaires de validité concomitante avec les attitudes envers l'inclusion ainsi que d'invariance de mesure selon le genre ont été observées. Cette validation psychométrique initiale, réalisée dans un contexte franco-canadien, répond à une lacune de mesure dans la littérature en EPS. Ces résultats préliminaires témoignent du potentiel de l'instrument, tout en soulignant la nécessité de poursuivre sa validation dans des contextes diversifiés, au moyen de devis longitudinaux et d'études de validité prédictive.

Mots-clés**:** inclusion; pratiques d'enseignement; hétérogénéité; questionnaire autorapporté; développement d'échelle; compétence

Introduction

Global perspectives and policies guide educational environments towards universality, accessibility, and equity within educational systems (UNESCO, 2019). These frameworks emphasize enhancing learning and social participation for all students, making this mission particularly crucial in PE (Block & Obrusnikova, 2007).

Differentiated instruction (DI) is one answer to inclusive education (Tomlinson, 2014). Graham et al. (2021) defined DI as the "use of proactive planning and inclusive practices to create accessible learning experiences to meet the needs of all learners in heterogeneous classrooms" (p. 164). As a complex construct, DI actions refer to a set of teaching practices that rely on the teachers' intention to act and their abilities to cope with the diversity of students (Tomlinson, & Imbeau, 2023).

Physical education (PE) teachers face considerable challenges in implementing inclusive practices (Haegele et al., 2020): lack of training, negative prior experiences with students having special educational needs (SEN), time constraints, and misconceptions about inclusive practices (Hutzler et al., 2019; Tarantino et al., 2022). Nevertheless, one major challenge in implementing DI lies in making its definition operational and measurable for teachers (Graham et al., 2021).

Purpose and Classification of Differentiated Instruction Measurement Instruments

Self-report questionnaires measuring DI practices serve distinct purposes that shape their development and validation requirements (Streiner et al., 2024). Instruments may be designed to distinguish between individuals or groups at a single point in time (discriminative purpose), examining variation in current practice implementation, or to monitor change over time (evaluative purpose), tracking teachers' development across professional learning experiences (Boateng et al., 2018; DeVellis & Thorpe, 2021)

Teacher practices can be measured through observational tools or self-report instruments. Observational tools involve external raters coding behaviors using structured protocols, providing objective, contextualized data but requiring substantial resources including observer training and multiple observation sessions (Bell et al., 2012; Hill et al., 2012). Self-report instruments ask teachers to reflect on their own practices, beliefs, or competencies. While susceptible to social desirability bias, self-reports are feasible for large-scale data collection, capture teachers' perspectives on practices that may not be visible during single observations and represent the most pragmatic methodological choice for initial validation studies examining general practice patterns and perceived competencies (Kyriazos & Stalikas, 2018; Podsakoff et al., 2012).

Beyond research applications, self-assessment tools enable teachers to identify practices requiring development, functioning as diagnostic guides for individualized learning pathways while maintaining scientific rigor (Prast et al., 2015; Van Geel et al., 2022).

Measures of Differentiated Instruction Practices

A review of recent literature examines existing DI measurement instruments, highlighting critical gaps justifying development of a PE-specific tool. Six self-reported questionnaires and one observation scale were identified (see Appendix A)

The *Differentiated Instruction Scale* (DIS) by Roy et al. (2013) is a validated French-Canadian language instrument. This 12-item scale measures frequency of DI practices among 125 primary teachers using a five-point Likert scale. Structural validity confirmed a two-factor

structure (instructional adaptation and progress monitoring) with acceptable psychometric properties.

The *Differentiation Self-Assessment Questionnaire* (DSAQ) by Prast et al. (2015) comprises 56 items organised into five theoretical scales using a five-point Likert scale. Validation with 268 mathematics teachers from Netherlands showed superior fit for a two-factor structure over the theoretical five-factor model. Van Geel et al. (2022) adapted this instrument (DSAQ+) revealing lower implementation frequencies for practices addressing student heterogeneity.

The *Differentiated Instruction Questionnaire* (DI-QUEST) by Coubergs et al. (2017) demonstrated strong cross-cultural validity with 1,574 Belgian teachers from kindergarten to secondary. This 31-item instrument encompasses five factors: teachers' mindset, ethical compass, flexible grouping, assessment for learning, and adaptive teaching using a seven-point Likert scale. Subsequent validation studies in 431 primary and secondary teachers from Hong Kong (Yuen et al., 2022), and 1935 secondary teachers from China (Meijie et al., 2023) confirmed that flexible grouping and assessment practices predict DI implementation. Wen and Cai's (2024) PE adaptation of DI-QUEST with 527 Chinese preservice teachers showed that growth mindset demonstrated the strongest effect on DI implementation, suggesting PE contexts may emphasise different factors than general education settings.

The *Lieberman-Brian Inclusion Rating Scale for Physical Education* (LIRSPE) by Lieberman et al. (2019) represents the only instrument specifically designed for PE contexts. This observational scale evaluates inclusive environment creation with acceptable validity.

Role of Teacher Beliefs in Differentiated Instruction Implementation

The theoretical relationship between attitudes, beliefs, and practice can be understood through Ajzen's Theory of Planned Behaviour (Ajzen, 1991, 2012), which posits that individual behaviour is determined by intentions shaped by attitudes toward the behaviour, subjective norms, and perceived behavioural control. The review by Knauder and Koschmieder (2019) underscores that competency beliefs, perceived self-efficacy, and attitudes toward inclusive education represent critical factors impacting teachers' inclusive practices, aligning with the Ajzen framework. These findings underscore the theoretical mechanism wherein individual teachers' attitudes and self-efficacy beliefs shape their intentions and subsequent practices: competency perceptions function as proximal determinants of teachers' willingness and capacity to implement DI strategies at the individual practitioner level (Ajzen, 1991, 2012).

According to Van Geel et al. (2019), previous works on DI have provided only limited understanding of the complex interaction between teachers' inclusive practices and beliefs. This limitation is particularly pronounced in PE research, where studies have predominantly examined inclusion through the narrow lens of either teachers' beliefs or students' perspectives, without systematically connecting these domains (Haegele and Sutherland, 2015; Tant and Watelain, 2016). De Neve et al. (2015) documented significant correlations between perceived self-efficacy and DI implementation, while Hutzler et al. (2019) revealed substantial associations between attitudes, normative beliefs, control beliefs, and inclusive practices specifically in PE contexts.

Conceptual Framework

This study integrates Tomlinson's (2014) DI theoretical model with operational frameworks from Prast et al. (2015) and Van Geel et al. (2022) to bridge theory-practice gaps. The resulting framework conceptualises DI practices within three interconnected dimensions: (1) student heterogeneity, (2) instructional adaptation, and (3) progress monitoring and evaluation.

The first dimension, *considering student heterogeneity*, involves how teachers systematically relate to readiness, interests, preferences, or other variables of diversity (Tomlinson & Imbeau, 2023). Prast et al.'s (2015) set this dimension in two differentiation steps: identifying educational needs and formulating differentiated goals considering student's needs priorities.

The second dimension, *instructional adaptation*, requires teachers to adjust their practices both to meet students' needs, enabling learners to reach their zone of proximal development (Tomlinson & Imbeau, 2023) and fostering their motivation (Wilhelmsen et al., 2019). Tomlinson (2014) operationalizes this dimension through four components: adjusting content (learning goals, knowledge, and competencies), teaching and learning processes (inclusive and cooperative teaching models, scaffolding processes) learning environment (flexible time management, flexible grouping), or products (various assessment methods for students to demonstrate their competencies). Prast et al. (2015) complement this dimension by grouping instruction practices (whole-class, subgroup, or individual adaptations) and by various learning tasks practices.

The third dimension, *progress monitoring and evaluation*, emphasises the iterative nature of DI implementation through continuous monitoring and regulation of student learning via formative assessment and systematic progress tracking. This involves concrete evidence (performance indicators, observations, assessment results) informing cyclical decision-making for DI implementation (Prast et al., 2015; Van Geel et al., 2022).

Critical Analysis and Justification for Questionnaire's Development

Four major limitations justify developing a questionnaire of DI practices in PE (Q-DIPPE). First, existing instruments address general education or mathematics contexts. Only two studies examined PE: Lieberman et al. developed an observation tool rather than self-report questionnaire, while Wen and Cai (2024) focused on Chinese preservice PE teachers.

Second, substantial language and cultural gaps exist. Roy et al.'s (2013) French-validated questionnaire targets elementary instruction, not PE. Most instruments are validated in English or Chinese contexts, revealing a critical gap for French-Canadian speaking PE teachers. Research indicates instruments must be developed with target populations and tailored to specific settings (Desbiens et al., 2018), necessitating contextual PE-specific DI measurement.

Third, most instruments assess only practice frequency. Van Geel et al. (2022) emphasised measuring both frequency and perceived competency, yet no instrument captures this dual perspective in PE.

Fourth, policy alignment limitations emerge regarding inclusive education. Some policies conceptualise DI as encompassing universal practices, subgroup adaptations, and individual modifications (Ministère de l'Éducation du Québec, 2021), yet instruments inadequately represent this tiered perspective.

Purposes

This study aims to develop and provide an initial psychometric validation of a self-report questionnaire measuring DI practice frequency and perceived competency in PE. The instrument development is grounded in an integrated theoretical framework combining Tomlinson's (2014) model with the procedural elements of Prast et al. (2015) and Van Geel et al. (2022), encompassing three-dimensions: (1) assessment of student heterogeneity, (2) instructional adaptation, and (3) progress monitoring and evaluation.

A second objective is to describe PE teachers' self-reported frequency and perceived competency of DI practices and to identify the most and least frequently implemented practices.

Methods

Following Boateng et al.'s (2018) systematic framework, this study implemented the foundational phases of scale development and validation. While acknowledging that construct validation is a continuous and never-ending process (Boateng et al., 2018; Messick, 1995), the present study focuses on establishing the initial foundation through rigorous execution of three phases (items development, scale development, and scale evaluation). This approach aligns with recommendations that scale development should begin with comprehensive assessment of content validity, internal structure, and preliminary construct validity before broader validation across multiple contexts (DeVellis & Thorpe, 2021; Streiner et al., 2024).

Phase 1: Item Development, Content, and Face Validity

The first phase included item development, content, and face validity assessments (Boateng et al., 2018). A panel of four experts in inclusive education was assembled based on specific expertise criteria. The panel composition included: (1) a specialist in physical education (PE) with 15+ years of experience in inclusive PE practices, (2) an educational sciences researcher with expertise in differentiated instruction, (3) a psychometrician with experience in scale development for educational contexts, and (4) a special education specialist with knowledge of adaptive teaching strategies. This panel size aligns with recommendations for content validation of domain-specific instruments where expert diversity and specialized knowledge are prioritized over panel size alone (Grant & Davis, 1997).

First, experts engaged in a deductive phase leveraging existing theoretical foundations and validated instruments to ensure theoretical alignment with the three iterative dimensions of the DI conceptual framework (student heterogeneity; instructional adaptation, and progress monitoring). They also conducted content analysis of existing instruments examining domains including universal design principles, inclusive teaching strategies, and adaptive PE practices. Rather than developing entirely new items, experts strategically modified and contextualized the Roy et al. (2013) questionnaire for PE, adapting its 12 items across two validated subscales: pedagogical adaptations and monitoring of academic progress. This hybrid approach was selected to maintains theoretical alignment with established measurement frameworks, to build upon psychometrically sound items with demonstrated validity and reliability, and to ensures efficient contextualization to the PE domain (Streiner et al., 2024).

Following the deductive phase, the four experts conducted an inductive phase through an exploratory focus group discussion (Vogt, King, & King, 2004). The 2-hour session was facilitated by a trained moderator following established focus group protocols (Krueger & Casey, 2015), with discussions structured around the three conceptual dimensions. This focus group informed item refinement and generation of additional context-specific items for the PE domain (Vogt et al., 2004). The integration of deductive (expert panel, literature review, existing instruments) and inductive (focus group with practitioners) approaches represents a triangulated method for establishing content validity, which strengthens the representativeness and relevance of the final item pool (Morgado et al, 2017).

The questionnaire incorporates two distinct five-point Likert-type scales. The first scale assesses the frequency of DI practices, while the second measures self-efficacy beliefs. The choice of five-point scales is grounded in three considerations. First, it is supported by robust empirical evidence demonstrating that five-point scales offer optimal psychometric properties: recent

research establishes that scales with four to seven response options optimize reliability and validity (Abulela & Khalaf, 2024) with marginal psychometric gains beyond seven point (DeCastellarnau, 2018). Second, five-point scales strike an optimal balance between cognitive demands placed on respondents and response differentiation capacity, thereby reducing survey fatigue while maintaining measurement precision (Abulela & Khalaf, 2024). Third, it ensures methodological consistency with Roy et al.'s (2013) questionnaire from which items were adapted, thereby facilitating cross-study comparisons.

All response options were accompanied by verbal descriptors to enhance interpretability and measurement consistency. Contemporary research confirms that fully labelled scales (all-points defined) yield superior test-retest reliability and reduce response bias compared to end-point-only labelling (Menold, 2020; Moors et al., 2014). For the implementation frequency scale, we employed frequency-based anchors: (1) *Never*, (2) *Rarely*, (3) *Sometimes*, (4) *Often*, and (5) *Always*. For the self-competency scale, agreement-based anchors were used: (1) *Strongly Disagree*, (2) *Disagree*, (3) *Neither Agree nor Disagree*, (4) *Agree*, and (5) *Strongly Agree*. These descriptors align with established psychometric conventions that distinguish between unipolar intensity measures and bipolar directional measures (Abulela & Khalaf, 2024) and were selected based on their clarity, familiarity to respondents, and empirically demonstrated capacity to approximate interval-level measurement properties when items are aggregated (Kyriazos & Stalikas, 2018).

Content and Face Validity Assessment

Following Boateng et al.'s (2018) recommendation, initial item generation produced 37 items in French language (see Appendix B). Content validity was assessed through the expert panel discussions, a recognized approach for establishing content validity in scale development (Boateng et al., 2018). This method enables experts to collaboratively refine relevance, clarity and representativeness, and achieve consensus through iterative dialogue (Krueger & Casey, 2015). This process is particularly valuable for developing context-specific instruments where nuanced understanding of the domain is essential (Morgado et al., 2017).

Face validity testing involved cognitive group interviews with six educational advisers (20+ years PE experience) to assess item appropriateness to targeted constructs (Streiner et al., 2024), leading to word clarification and example provision. Finally, three PE teachers (15-22 years' experience) completed the questionnaire and participated in individual cognitive interviews with the principal author to verify item intelligibility, identify administration difficulties, and minimise measurement errors.

Phase 2: Scale Development

The second phase consisted of scale construction, including administering the questionnaire, reducing the number of items, and exploring the factor structure.

Recruitment Procedure and Participants

The ethics committee of the principal researcher's university approved all procedures. A non-probabilistic convenience sampling approach was employed, leveraging the proximal effect to facilitate participant recruitment. This approach was justified on pragmatic grounds (Clarke & Visser, 2019). Consistent with the exploratory purpose of the instrument validation study, emphasis was placed on assessing psychometric properties rather than on population parameter estimation (Kyriazos & Stalikas, 2018).

After obtaining approval, all school boards in the province of Quebec ($N = 60$) were invited to participate by mail, representing approximately 5,000 PE teachers. The final sample includes 240 teachers (M_{age} = 41.5, SD = 9.0; 130 women and 110 men) from various backgrounds (50 school boards; 226 public schools and 14 private schools), gender (54% women; 46 % men), levels (72% preschool-primary; 28% secondary), and experience (M_{exp} = 15.5; SD = 8.4; novice 17%; middle 48%; experienced 35%; range 1-37 years). Table 1 presents the main sociodemographic characteristics.

The final sample size exceeded established thresholds for psychometric validation studies. According to guidelines (Kline, 2016), a sample of n = 200 represents the minimum adequate threshold for factor analysis, while sample sizes between 200-300 are classified as "fair" to "good" for establishing stable factor solutions (Comrey & Lee, 1992). Furthermore, the sample satisfied recommended participant-to-item ratios of 5:1 to 10:1 (Hair et al., 2010). To mitigate sampling bias, the final sample demonstrated adequate diversity across key demographic variables including gender, teaching level, years of experience, and institutional affiliation (school boards; public and private schools), thereby enhancing the representativeness of the sample within the constraints of the sampling method employed (Turner, 2020).

Table 1

Characteristics of Participants (N = 240)

	Frequency	%
Gender		
Women	130	54.2
Men	110	45.8
Teaching level		
Primary	173	72.1
Secondary	67	27.9
Public school board	226	94.2
Teaching qualification		
Bachelor in PE	203	84.6
Masters	24	10.0
Other	13	5.4
Teaching experience	15.5	8.4
Novice (0-6 years)	41	17.1
Middle (7-18 years)	114	47.5
Experienced (19-35 years)	85	35.4
At least of SEN student in their group	203	82.9
Having participated to IEP meeting	80	32.7

Phase 2 Data Analyses

Data were analysed using SPSS 29. All the missing values on the independent variables (.008%) were filled in with the average score of the scales for continuous variables or the score for the most common category for categorical variables (Tabachnick et Fidell, 2019). No missing data

was observed for the dependent variables. Prior to conducting the analyses, the assumptions of null hypothesis significance testing were examined. Tests for normality (Shapiro-Wilk) and homogeneity of variance (Levene's test) confirmed that the data met the required assumptions for parametric testing. When Levene's tests for homogeneity of variance were significant ($p < .05$), Welch's correction for unequal variances was applied.

All items displayed kurtosis values between -7 and 7, and many displayed coefficients of skewness over recommended values (between -3 and 3) (Kline, 2016). To determine the optimal number of factors, total variance explained (Tinsley & Tinsley, 1987), eigenvalue, scree plot inspection method (Cattell, 1966) and Minimum Average Partial method (MAP; Costello & Osborne, 2005; Izquierdo et al., 2014; Velicer, 1976) were conjointly used. Exploratory Factor Analysis (EFA) with maximum likelihood factors extraction and oblimin rotation were conducted (Gaskin & Happell, 2014). This rotation was chosen because theoretical dimensions were expected to be correlated (Field, 2017) as it was the case in previous scales development (Prast et al., 2015). Item deletion was made according to specific conditions: items without coefficients greater than .32 on factors were deleted, and items with cross-loadings greater than .32 were deleted (Osborne et al., 2008). The PCA solution was deemed acceptable when three indices reached acceptable values (Dziuban & Shirkey, 1974): Bartlett's test ($p < 0.05$); Kaiser-Meyer-Olkin test (KMO > .80); and total variance explained (> 50%; Tinsley & Tinsley, 1987). Internal consistency was tested using Cronbach's alpha and McDonald's Omega with each subscale of the frequency and competence instrument separately. According to Nunally (1978), reaching values over .70 is recommended for basic research designs.

Descriptive analyses (mean scores, standard deviations) were conducted for each scale, subscale, and item. Following Roy et al. (2013) and Prast et al. (2015), items one standard deviation above the total factor mean were considered most frequently used, while one standard deviation below were considered least frequently used.

Phase 3: Scale Evaluation

Using the same sample, this phase was designed to explore initial scale reliability and validity. Concurrent validity was examined by computing correlations between each of the four subscales of the questionnaire developed in this study and theoretically related constructs measured with other instruments, namely attitudes towards inclusion and willingness to teach in inclusive PE.

Measures of Theoretically Related Variables

In addition to the Q-DIPPE, teachers completed an adaptation of the *Multidimensional Attitudes Toward Inclusive Education Scale* (Mahat, 2008). As established by Mahat (2008), the questionnaire measures teachers' attitudes towards school inclusion of students with special needs using cognitive (6 items), affective (6 items), and behavioural dimensions (6 items).

Willingness to teach was measured using an adaptation of the *Teachers' Willingness to Work with Severe Disabilities Scale* (Rakap & Kaczmarek, 2010). The instrument includes four vignettes and eight items asking to what extent teachers would accommodate students with special needs (behavioural, intellectual, or physical difficulties and sports giftedness). A Cronbach's alpha of .94 was reported by the authors and replicated by MacFarlane and Woolfson (2013).

Phase 3 Data Analysis

Confirmatory factor analysis with maximum likelihood robust estimator was used to verify scale structure. Four fit indices evaluated model fit: Root Mean Square Error of Approximation (RMSEA) should be under .060 (Schmitt, 2011), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) should be over .900 (Hu & Bentler, 1999), and Root Mean Square of Residuals (RMSR) should be under .080 (Browne & Cudek, 1993). Multiple-group analysis verified whether the four-factor structure was consistent across gender groups.

Confirmatory factor analysis tested configural, metric, and scalar invariance. Configural invariance tests whether factor structure is the same across groups. Metric invariance verifies whether factor loadings vary across groups. Scalar invariance examines whether item intercepts vary across groups. Models were compared according to CFI, TLI, and RMSEA differences: changes less than 0.01 indicated model invariance (Chen, 2007). Analyses were conducted using Amos 29. Concurrent validity was tested using Pearson's correlations in SPSS 29. Associations were analysed between each of the four subscales of the DI practices questionnaire, in terms of frequency and competency, and attitudes (cognitive, affective, behavioural) and willingness to teach (behavioural difficulties, intellectual difficulties, physical difficulties, sports giftedness) variables. Correlation coefficients under .30 were considered small effects, between .30 and .50, moderate effects, and higher than .50, large effects (Cohen, 2013).

Results

Results indicate that Q-DIPPE meets the initial stages of psychometric validation for a French-Canadian speaking PE context.

Results Phase 1: Item Development, Content and Face Validity

Experts generated 37 items based on theoretical models and previous research using deductive methods, content analysis, and focus groups. Items were refined through consensus and validated for clarity by three PE teachers via cognitive interviews (see Appendix A).

Results Phase 2: Scale Development

In phase 2, principal component analyses were conducted with the frequency 37-item scale. The analysis suggested a 16-item solution scale divided into four factors: (1) considering student heterogeneity (4 items); (2) adjusting practices to support autonomy and motivation (4 items); (3) adjusting practices to handle heterogeneity (4 items); and (4) monitoring progress of students with needs (4 items). There were no cross-loadings between factors, and items loaded highly on their respective factors. The Kaiser-Meyer-Olkin value was adequate (.827), and Bartlett's test was significant ($\chi^2_{(120)} = 993.17$, $p < .001$). The explained variance was 55.19%.

The same four-factor model was tested with the competency scale and displayed adequate fit. The Kaiser-Meyer-Olkin value was satisfactory (.877), Bartlett's test was significant ($\chi^2_{(120)} = 1148.95$, $p < .001$) and the explained variance was 57.75%. However, one item ('propose tasks based on students' interests') from the first factor (considering heterogeneity) in the 'frequency' model loaded on the third factor (adjusting practices to handle heterogeneity) in the 'competency' model (Table 2). To maintain theoretical consistency, the model obtained with the frequency scale was retained for subsequent analyses, with increased vigilance in interpreting future results for this item. According to internal consistency values, the first factor provided poorer reliability (frequency: $a = .54$; $\Omega = .54$; competency: $a = .61$; $\Omega = .63$) than the three other factors in both

models, whereas it did not reach the recommended value (Nunnally, 1978). In the 'frequency' model, the third factor was also under, but closer, to this recommended value (a = .67; Ω = .67).

Table 2

Loading Factors and Internal Consistencies

	Frequency				Competency			
	I	II	III	IV	I	II	III	IV
Factor 1 Considering the heterogeneity		a = .54; Ω = .54				a = .61; Ω = .63		
Item 1 *Determine tasks based on students' interests*	.68						.77	
Item 4 Determine tasks based on students' gender	.68				.49			
Item 13 Determine tasks based on students' preferences	.45				.69			
Item 15 Determine tasks based on students' cultural background	.67				.75			
Factor 2 Practices /autonomy and motivation		a = .72; Ω = .72				a = .73; Ω = .73		
Item 16 Use different methods of presentation			-.76				-.74	
Item 28 Varied modalities for assessment tasks			-.86				-.79	
Item 29 Provide a range of learning challenges allowing student to select the one that best meets their needs			-.63				-.61	
Item 33 Facilitate students' autonomous access to the resources and materials for assignment completion			-.62				-.71	
Factor 3 Practices /handle heterogeneity		a = .67; Ω = .67				a = .70; Ω = .70		
Item 2 Adjust the tasks' parameters to accommodate for the differences among students		.76				.73		
Item 6 Adjust the workload to the students' skills		.67				.53		
Item 23 Adjust a task's difficulty or intensity to accommodate students' needs		.67				.63		
Item 24 Adjust the assessment task based on the observations of the group's progress		.55				.52		
Factor 4 Monitoring progress/ special needs		a = .83; Ω = .83				a = .81; Ω = .81		
Item 18 Monitor the progress of students who require special learning support				-.78				-.79
Item 25 Adjust one student's assessment task based on his progress				-.74				-.75
Item 30 Propose distinct learning tasks for students having special educational needs.				-.83				-.77
Item 35 Evaluate the efficacy of the interventions provided to students with special educational needs.				-.79				-.79
Explained variance		55.19%				57.75%		

Note. This is a free translation by the authors. English items have not been validated. French validated items are available in Appendix A; To facilitate the reading, all values < .32 were not reported.; Item in *italic* do not load on the same factor in the frequency and in the competence models. a = Cronbach's alpha; Ω = McDonald's Omega

Table 3 presents descriptive analyses for each subscale and item. Mean and standard deviation scores for the frequency scale were 3.38 (SD = 0.54), range [2.41-4.05], while those for the competency scale were 3.54 ± 0.56, range [2.76-4.20]. Mean scores indicate moderate to high self-reported frequency and perceived competency. The three most frequently reported items (2, 6, 23) focused on DI practices that enable handling heterogeneity by adjusting content

(frequency and competency item 2; frequency items 6-23). Three items (frequency and competency item 15; frequency item 30) scored at least one standard deviation below the total mean score of their respective factors.

Table 3

Means, Standard Deviations, and Median Scores for Frequency and Competency

Subscales Items	F Mean	F *SD*	F Median	C Mean	C *SD*	C Median
Scale total mean score	3.38	.54	3.38	3.54	.56	3.53
Considering heterogeneity	**3.16**	**.77**	**3.25**	**3.43**	**.75**	**3.50**
Item 1 Determine tasks based on interests	3.86	.88	4.00	4.17	.70	4.00
Item 4 (…) gender	3.33	1.12	3.00	3.64	.93	4.00
Item 13 (…) preferences	3.05	1.13	3.00	3.15	1.10	3.00
Item 15 (…) cultural background	2.41[1]	1.59	3.00	2.76[1]	1.56	3.00
Adjusting practices/autonomy and motivation	**3.35**	**.80**	**3.50**	**3.59**	**.74**	**3.75**
Item 16 Use different methods of presentation	3.64	1.01	4.00	3.84	.96	4.00
Item 28 Varied modalities for assessment tasks	3.09	1.01	3.00	3.41	1.01	4.00
Item 29 Provide a range of challenges (…)	3.23	1.14	3.00	3.50	1.02	4.00
Item 33 Facilitate autonomous access (…)	3.45	1.01	4.00	3.60	1.03	4.00
Adjusting practices/heterogeneity	**4.01**	**.62**	**4.00**	**4.01**	**.61**	**4.00**
Item 2 Adjust the task's parameters (…)	4.22[2]	.87	4.00	4.20[2]	.85	4.00
Item 6 Adjust the workload (…)	4.05[2]	.89	4.00	4.04	.82	4.00
Item 23 Adjust a task's difficulty or intensity (…)	3.97[2]	.82	4.00	4.00	.77	4.00
Item 24 Adjust the assessment task (…)	3.80	.93	4.00	3.81	.93	4.00
Monitoring the progress/special needs	**3.02**	**.88**	**3.00**	**3.13**	**.83**	**3.25**
Item 18 Monitor the progress (…)	2.91	1.06	3.00	3.10	1.02	3.00
Item 25 Adjusting assessment task (…)	3.31	1.09	3.00	3.28	1.02	3.00
Item 30 Propose distinct tasks (…)	2.84[1]	1.10	3.00	3.05	1.05	3.00
Item 35 Evaluate efficacy of interventions (…)	3.01	1.08	3.00	3.10	1.05	3.00

Note. F= Frequency; C= Competency; [1]items having one standard deviation under total mean score; [2]items having one standard deviation over total mean score; Ranges between minimum and maximum scores were 1 to 5 for all items in both scales.

Results Phase 3: Scale Evaluation

Results from phase 3 indicated good fit of the CFA model with four factors for the frequency (χ^2 = 144.28, *df* = 98, *p* = .002, TLI = .94, CFI = .95, RMSEA = .04, LO 90% = .03, HI 90% = .06; RMSR = .05) and competency (χ^2 = 153.21, *df* = 98, *p* = .000, TLI = .94, CFI = .95, RMSEA = .05, LO 90% = .03, HI 90% = .06; RMSR = .05) models. Standardised factor loadings are presented in Table 4. In the 'frequency' model, standardised factor loadings varied between .34 and .82. The two lower factor loadings (< .40) were on the same factor. In the 'competency' model, standardised factor loadings varied between .43 and .80. The lowest factor loading was on the same factor as in the 'frequency' model (Table 4).

Table 4

Parameter Estimates (standardized solution) for Each Model

	Model 1 Frequency	Model 2 Competency
Regression coefficients		
CH ® item 1	.36	.62
CH ® item 4	.51	.61
CH ® item 13	.70	.53
CH ® item 15	.34	.43
PH ® item 2	.52	.58
PH ® item 6	.51	.58
PH ® item 23	.65	.66
PH ® item 24	.64	.63
PAM ® item 16	.65	.16
PAM ® item 28	.71	.28
PAM ® item 29	.64	.29
PAM ® item 33	.51	.33
MP ® item 18	.64	.64
MP ® item 25	.72	.69
MP ® item 30	.77	.74
MP ® item 35	.82	.80
Covariances		
CH ↔ PH	.64	.92
CH ↔ PAM	.46	.58
CH ↔ MP	.44	.64
PH ↔ PAM	.48	.70
PH ↔ MP	.60	.66
PAM ↔ MP	.44	.50

Note. All relations are significant $p < .05$; CH = Considering heterogeneity; PH = Adjusting practices for heterogeneity; PAM = Adjusting practices for autonomy and motivation; MP = Monitoring progress for students having special educational needs.

In both frequency and competency models, CFI, TLI, and RMSEA differences were ≤ .01, indicating model invariance across gender groups. All factors were significantly positively correlated (within and between scales) (Table 5). Correlations ≥ .79 between corresponding frequency and competency factors indicated strong associations between perceived competency and implementation frequency.

Table 5

Intercorrelations Factors for Both Scales and Descriptive Statistics

	1	2	3	4	5	6	7	8
1. F_Considering heterogeneity	-	.26**	.34**	.25**	.72**	.21**	.26**	.24**
2. F_Practices/autonomy		-	.32**	.38**	.27**	.79**	.33**	.35**
3. F_Practices/heterogeneity			-	.44**	.44*	.35**	.81**	.46**
4. F_Monitoring progress				-	.26**	.28**	.38**	.80**
5. C_Considering heterogeneity					-	.40**	.54**	.41**
6. C_Practices/autonomy						-	.50**	.39**
7. C_Practices/heterogeneity							-	.50**
8. C_Monitoring progress								-

Mean	3.16	3.35	4.01	3.02	3.43	3.59	4.01	3.13
(*SD*)	(.77)	(.80)	(.62)	(.88)	(.75)	(.74)	(.61)	(.83)
Skewness	0.09	-.16	-.58	-.28	.03	-.39	-.42	-.37
Kurtosis	-.32	-.40	.68	-.35	-.56	-.17	.39	-.10

Note. ** $p < .01$ bilateral; F = Frequency; C = Competency.

For concurrent validity, all four factors of both frequency and competency subscales were associated (≤ .30) with PE teachers' attitudes and willingness towards inclusion (Table 6 and Table 7).

Table 6

Frequency Factors Correlations with Attitudes and Willingness Toward Inclusion

	Att total	Att Cogn	Att Beh	Att Aff	Beh diff	Phys diff	Int diff	Gifted
Cons/heterogeneity	.04	-.03	.08	.06	.14*	.09	.15*	.18**
Practices/autonomy	.22**	.16*	.19**	.19**	.27**	.25**	.31**	.24**
Practices/heterogeneity	.16*	.13*	.18*	.09	.22**	.22**	.21**	.19**
Monitoring progress	.26**	.22**	.21**	.22**	.21**	.22**	.27**	.27**

Note. * $p < 0{,}05$. ** $p < 0{,}01$. *** $p < 0{,}001$; Att = Attitudes; Cogn = Cognitive; Beh = Behavioural; Aff = Affective; Beh diff = Behavioural difficulties; Phys diff = Physical difficulties; Int diff = Intellectual difficulties;

However, when examining the correlations in Table 7, the first factor of the 'frequency' scale showed no significant associations with any of the attitude dimensions, indicating that teachers' reported frequency of considering student heterogeneity was not related to their overall attitudes toward inclusion, cognitive, behavioral, or affective aspects. In contrast, for the 'competence' scale, the first factor was significantly correlated only with the behavioral dimension of attitudes ($r = .18$, $p < .01$), suggesting that teachers' perceived competence in considering heterogeneity relates specifically to their willingness to act inclusively, rather than to their overall attitudes or affective/cognitive components. Additionally, other factors of the competency scale, such as practices/autonomy and monitoring progress, showed broader associations with multiple attitude dimensions, highlighting the differential links between specific subscales and attitudes toward inclusion.

Table 7

Competency Factors Correlations with Attitudes and Willingness Toward Inclusion

	Att total	Att cogn	Att Beh	Att Aff	Beh diff	Phys diff	Int diff	Gifted
Consi/heterogeneity	.09	.04	.13*	.08	.18**	.18**	.23**	.24**
Practices/autonomy	.21**	.16**	.22**	.16*	.26**	.30**	.31**	.24**
Practices/heterogeneity	.16*	.14*	.18**	.09	.18**	.18**	.21**	.13*
Monitoring progress	.24**	.19**	.17**	.23**	.16*	.16*	.22*	.25*

Note. * $p < 0{,}05$. ** $p < 0{,}01$. *** $p < 0{,}001$; Att = Attitudes; Cogn = Cognitive; Beh = Behavioural; Aff = Affective; Beh diff = Behavioural difficulties; Phys diff = Physical difficulties; Int diff = Intellectual difficulties.

Discussion

This study validated the Q-DIPPE using a three-phased approach to instrument development following Boateng et al.'s (2018) systematic framework. First phase produced 37 items through a triangulated approach combining deductive expert panel analysis, inductive focus group discussions, and face validity testing with educational advisers and PE teachers, establishing initial content validity. Phase 2 reduced the instrument to 16 items across four factors through exploratory factor analysis and Phase 3 confirmed the four-factor structure through confirmatory factor analysis with good model fit indices.

Four-Factor Structure: Psychometric Evidence and Practice Patterns

Results of Phase 2 demonstrated acceptable structure validity (KMO > .82, variance > 55%) for both scales, with internal consistencies ranging from α = .54-.83 and moderate to high self-reported implementation scores (M = 3.38-3.54).

This four-factor structure respects the theoretical constructs. The first factor includes practices related to heterogeneity. Factor 2 covers practices supporting students' autonomy and responsibility. This factor brings new insight reflecting how DI practices can sustain students' motivation in PE (Girard et al., 2023). Moreover factor 3 encompasses practices that address heterogeneity through content or process adjustments and converges with previous scales, such as the *Adaptive Instruction* scale from DI-QUEST (Coubergs et al., 2017). The fourth factor concerns monitoring students' progress. Those factors bring an important contribution that clarifies the conceptual dimension of DI.

Psychometric Considerations and Literature Comparison

One item, *'propose tasks based on students' interests'*, did not load on the same factor in the frequency and competency scales, reflecting challenges commonly observed in educational practice measurement. Indeed, DI measurement faces unique challenges due to complex practices spanning multiple domains, with cultural and contextual factors influencing implementation (Meijie et al., 2023). The low internal consistency values particularly for factor 1 in the frequency scale (α = .54, Ω = .54) and to a lesser extent factor 3 (α = .67, Ω = .67), align with systematic patterns identified across DI instruments. Prast et al. (2015) reported similar challenges with their five-factor DSAQ model. Coubergs et al. (2017) acknowledged measurement difficulties with DI-QUEST ethical compass scale dimensions. More recently, Pereira et al. (2021) reported weaker construct-related validity for low-achieving students, highlighting the inherent complexity of capturing differentiated practices across diverse populations. This may reflect authentic teaching complexity rather than instrument inadequacy, as teachers may develop distinct competencies independently (Vita, 2023; Wilkinson & Penney, 2023).

Physical Education Context and Practice Patterns

Descriptive results showed that PE teachers use DI practices and feel relatively competent, potentially distinguishing them positively from other disciplines where implementation challenges are more pronounced (Hutzler, 2019; Van Geel et al., 2022). Content adjustments to address heterogeneity received highest frequency and competency scores, strongly supporting Whipp et al. (2014) findings that content-related DI practices predominate in PE contexts. This pattern contrasts with general education settings where teaching process-related practices typically receive

greater emphasis (Prast et al., 2015), likely reflecting PE's unique capacity for offering multiple options and skill levels within single lessons.

Notably, only targeted practices for students with special educational needs showed significantly lower implementation scores, contrasting with broader DI literature where targeted interventions consistently represent the most challenging tier (Massé et al., 2020; Van Geel et al., 2012). However, the averages scores in the middle-upper bracket suggests that PE teachers in the current sample demonstrate relatively advanced DI skills. However, lower scores for cultural responsiveness practices align with broader concerns about preparing PE teachers for increasingly diverse student populations (Vita, 2023), highlighting priority areas for future professional development.

Confirmatory Factor Analysis and Concurrent Validity Evidence

Phase 3 confirmed the four-factor structure through confirmatory factor analysis with good model fit indices (CFI = .95, TLI = .94, RMSEA = .04-.05), established measurement invariance across gender, demonstrated strong correlations between corresponding frequency and competency factors ($r \geq .79$), and provided preliminary evidence of concurrent validity through small but significant correlations ($r \leq .30$) with attitudes toward inclusion and willingness to teach diverse learners.

Dual-Scale Innovation and Validity Evidence

The dual-scale approach of the Q-DIPPE addresses critical gaps by measuring both frequency and competency, representing distinct constructs with different predictive relationships (Van Geel et al., 2022). This directly addresses limitations of previous instruments that failed to capture the complexity of belief-practice relationships (Coubergs et al., 2017; Prast et al., 2015).

Results revealed some small but significant correlations between competency and frequency (r = .20-.40), consistent with broader patterns in educational practice measurement (Klassen & Chiu, 2010; Van Geel et al., 2022). Correlations were expected given DI complexity and multiple influencing variables (Hutzler et al., 2019; Tarantino et al., 2022). These modest correlations reflect the authentic complexity of translating competency beliefs into consistent practice implementation within unpredictable PE settings.

Taken together, the findings support the concurrent validity of the scale, as the four factors of both frequency and competency subscales showed weak but significant associations ($r \leq .30$) with PE teachers' attitudes toward inclusion and willingness to include students. Nevertheless, the pattern of relationships varied across factors. In particular, the first factor of the frequency subscale was unrelated to any dimension of attitudes, whereas the corresponding factor of the competency subscale was linked exclusively to the behavioral component. This pattern suggests that frequently reporting attention to student heterogeneity does not translate into more positive inclusive attitudes. In contrast, perceived competence appears to play a more targeted role, especially in shaping teachers' behavioural intentions toward inclusive practice.

The Q-DIPPE's concurrent validity patterns mirror those across DI measurement studies, with weak to moderate correlations with attitudes towards inclusion (r = .15-.30; Coubergs et al., 2017; Yuen et al., 2022). This supports theoretical models proposing that effective DI requires specific pedagogical knowledge and contextual factors rather than general positive attitudes towards inclusion (Graham et al., 2021; Hutzler et al., 2019).

Theoretical Advances in Differentiated Instruction Measurement

The Q-DIPPE allows for capturing outcomes on teacher's universal practices (factors 1, 2, 3) while focusing on targeted practices for students with greater difficulties (factor 4), representing a significant advance over previous unidimensional approaches. First, this framework captures contemporary DI theory emphasizing proactive, student-centered approaches (Graham et al., 2021). Second, the Q-DIPPE uniquely incorporates motivation-supportive practices (Factor 2), addressing both cognitive and affective dimensions (Ryan & Deci, 2009). Finally, cultural responsiveness considerations distinguish this instrument from those developed in other cultural contexts (Tomlinson & Imbeau, 2023).

Limitations

This study has several limitations. First, reliance on self-reported measures introduces response bias, including dual-scale self-assessment challenges, potential discrepancies between self-assessed and actual competence, and recall bias for inconsistently implemented practices (Hogan, 2019; Podsakoff et al., 2012). While triangulation with observations would strengthen validity claims, self-reported measures remain the most feasible approach for initial large-sample validation (Hogan, 2019).

Second, convenience sampling and the 4% response rate (despite contacting all 5,000 estimated PE teachers in Quebec) introduce potential selection bias and limit generalizability (Morgado et al., 2017). The use of a single sample for both EFA and CFA is not ideal; however, the sample size (n = 240) exceeded recommended thresholds (Kyriazos & Stalikas, 2018), and this approach has precedent in comparable educational scale development studies (Pereira et al., 2021; Roy et al., 2013; Wen & Cai, 2024). Split-sample validation was impractical given the specialized population and recruitment constraints (Ferrando & Lorenzo-Seva, 2024).

Third, the five-point response scale may limit discrimination compared to seven-point scales, though this choice balanced response precision with respondent burden (DeCastellarnau, 2018). Fourth, test-retest reliability was not assessed due to practical constraints; temporal stability should be examined in future longitudinal studies (Boateng et al., 2018).

Finally, generalizability is constrained by the French-Canadian context. Cross-cultural validation with other French-speaking and anglophone populations would clarify whether the four-factor structure reflects universal or context-specific dimensions (Beni et al., 2017). These findings should be interpreted as exploratory until replicated.

Conclusion

This study establishes the initial psychometric foundation for the Q-DIPPE through Boateng et al.'s (2018) validation framework. It addresses critical gaps by providing the first validated measure specifically designed for PE that aligns with inclusive educational policy frameworks (UNESCO, 2019).

The Q-DIPPE identified four dimensions of DI practice and allows measure of both frequency and perceived competency. The instrument addresses the complex interaction between teacher beliefs and practice implementation identified by Van Geel et al. (2022). The Q-DIPPE incorporates a tiered conceptualization encompassing universal flexibility practices, targeted adaptations for subgroups, and intensive modifications for students with substantial challenges, reflecting evidence-based multi-tiered intervention models (Berkeley et al., 2009).

The questionnaire's conciseness and positive framing help minimize completion attrition (Boateng et al., 2018) The Q-DIPPE can be combined with other measures to investigate factors influencing DI practices or serve as a professional development tool by identifying practices requiring further implementation or competency development. While these findings encourage the instrument's use in French-Canadian PE contexts, it should be recognized that validation is an ongoing process requiring replication across diverse educational systems and populations. The instrument, although primarily validated, is critical for this domain and support forthcoming research.

Acknowledgments
The authors would like to thank the participants and the school boards for their collaboration. They also want to thank graduate students for their precious participation in the project.

References

Abulela, M. A. A., & Khalaf, M. A. (2024). Does the number of response categories impact validity evidence in self-report measures? A scoping review. *SAGE Open, 14*(1), Article 21582440241230363. https://doi.org/10.1177/21582440241230363

Alquraini, T., & Gut, D. (2012). Critical components of successful inclusion of students with severe disabilities: Literature review. *International Journal of Special Education*, *27*(1), 42–59.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2–3), 62–87. https://doi.org/10.1080/10627197.2012.715014

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of Response to Intervention: A snapshot of progress. *Journal of Learning Disabilities*, *42*(1), 85-95. https://doi.org/10.1177/0022219408326214

Block, M. E., & Obrusnikova, I. (2007). Inclusion in physical education: A review of the literature from 1995-2005. *Adapted Physical Activity Quarterly*, *24*(2), 103–124. https://doi.org/10.1123/apaq.24.2.103

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Clarke, V., & Visser, M. (2019). Pragmatic research methodology in education: Possibilities and pitfalls. *International Journal of Research & Method in Education, 42*(5), 455–469. https://doi.org/10.1080/1743727X.2018.1524866

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.).* Lawrence Erlbaum Associates.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 86–99. https://doi.org/10.4135/9781412995627.d8

Coubergs, C., Struyven, K., Vanthournout, G., & Engels, N. (2017). Measuring teachers' perceptions about differentiated instruction: The DI-Quest instrument and model. *Studies in Educational Evaluation*, *53*, 41–54. https://doi.org/10.1016/j.stueduc.2017.02.004

DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity, 52*(4), 1523–1559. https://doi.org/10.1007/s11135-017-0533-4

De Neve, D., Devos, G., & Tuytens, M. (2015). The importance of job resources and self-efficacy for beginning teachers' professional learning in differentiated instruction. *Teaching and Teacher Education*, *47*, 30–41. https://doi.org/10.1016/j.tate.2014.12.003

Desbiens, J.-F., Naila, B., Spallanzani, C., Vandercleyene, F., & Beaudoin, S. (2018). Validation d'un instrument pour mesurer les préoccupations d'enseignants stagiaires en ÉPS tunisiens [Validation of an instrument to measure the concerns of Tunisian PE teacher trainees]. *Anadolu University Journal of Education Faculty*, *2*(2), 158–177.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage Publications.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/h0036316

Ferrando, P. J., & Lorenzo-Seva, U. (2024). Determining sample size requirements in EFA solutions: A simple empirical proposal. *Multivariate Behavioral Research*, *59*(2), 285–299.

Field, A. (2017). *Discovering statistics using SPSS: North American edition*. Sage Publications.

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*(3), 511–521. https://doi.org/10.1016/j.ijnurstu.2013.10.005

Girard S, de Guise A-A, Hogue A-M and Desbiens J-F (2023) Changes in physical education teachers' beliefs regarding motivational strategies: A quasi-experimental study. *The Physical Educator* 80(6): 607-630. DOI: 10.18666/TPE-2023-V80-I6-11447

Graham, L. J., de Bruin, K., Lassig, C., & Spandagou, I. (2021). A scoping review of 20 years of research on differentiation: Investigating conceptualisation, characteristics, and methods used. *Review of Education*, *9*(1), 161–198. https://doi.org/10.1002/rev3.3238

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health, 20*(3), 269-274. https://doi.org/10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

Haegele, J. A., Hodge, S., Gutuskey, L., & Foley, J. T. (2020). Physical education teachers' perspectives on inclusion: A decade later. *Research Quarterly for Exercise and Sport*, *91*(4), 690–701. https://doi.org/10.1080/02701367.2019.1706297

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: Global edition*. Pearson Higher Education.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for their generalizability. *Educational Researcher, 41*(2), 56–64. https://doi.org/10.3102/0013189X12437203

Hogan, T. P. (2019). *Psychological testing: A practical introduction* (4th ed.). Wiley.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hutzler, Y., Meier, S., Reuker, S., & Zitomer, M. (2019). Attitudes and self-efficacy of physical education teachers toward inclusion of children with disabilities: A narrative review of international literature. *Physical Education and Sport Pedagogy*, *24*(4), 1–18. https://doi.org/10.1080/17408989.2019.1571183

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, *26*(3), 395–400. https://doi.org/10.7334/psicothema2013.349

King, D. W., King, L. A., & Vogt, D. S. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment*, *16*(3), 231-243. https://doi.org/10.1037/1040-3590.16.3.231

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741–756. https://doi.org/10.1037/a0019237

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

Knauder, H., & Koschmieder, C. (2019). Individualized student support in primary school teaching: A review of influencing factors using the Theory of Planned Behavior (TPB). *Teaching and Teacher Education*, *77*, 66–76. https://doi.org/10.1016/j.tate.2018.09.012

Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. *Psychology, 9(*11), 2531-2560. https://doi.org/10.4236/psych.2018.911145

Krueger, R. A., & Casey, M. A. (2015). Focus groups: A practical guide for applied research (5th ed.). Sage Publications.

Lieberman, L., Brian, A., & Grenier, M. (2019). The Lieberman-Brian Inclusion Rating Scale for Physical Education. *European Physical Education Review*, *25*(2), 341–354. https://doi.org/10.1177/1356336X17733595

MacFarlane, K., & Woolfson, L. M. (2013). Teacher attitudes and behavior toward the inclusion of children with social, emotional and behavioral difficulties in mainstream schools: An application of the theory of planned behavior. *Teaching and Teacher Education*, *29*, 46–52. https://doi.org/10.1016/j.tate.2012.08.006

Mahat, M. (2008). The development of a psychometrically-sound instrument to measure teachers' multidimensional attitudes toward inclusive education. *International Journal of Special Education*, *23*(1), 82–92.

Massé, L., Nadeau, M.-F., Verret, C., Gaudreau, N. et Lagacé-Leblanc, J. (2020). Facteurs influençant les attitudes des enseignant·e·s québécois·es envers l'intégration des élèves présentant des difficultés comportementales [Factors influencing the attitudes of Quebec teachers towards the integration of students with behavioral difficulties]. *Revue des Sciences de l'Éducation, 46*(1), 41–63. https://doi.org/10.7202/1070726ar

Meijie, B., Katrien, S., & Chang, Z. (2023). Variables that influence teachers' practice of differentiated instruction in Chinese classrooms: A study from teachers' perspectives. *Frontiers in Psychology*, *14*, Article 1124259. https://doi.org/10.3389/fpsyg.2023.1124259

Menold, N. (2020). Rating-scale labeling in online surveys: An experimental comparison of verbal and numeric rating scales with respect to measurement quality and respondents' cognitive processes. *Methods, Data, Analyses, 14*(2), 229-258. https://doi.org/10.12758/mda.2017.11

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*, 3. https://doi.org/10.1186/s41155-016-0057-1

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology*, *44*(1), 369-399. https://doi.org/10.1177/0081175013516114

Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best practices in quantitative methods. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 86–99). SAGE Publications. https://doi.org/10.4135/9781412995627

Pereira, N., Tay, J., Desmet, O., Maeda, Y., & Gentry, M. (2021). Validity evidence for the Revised Classroom Practices Survey: An instrument to measure teachers' differentiation practices. *Journal for the Education of the Gifted*, *44*(1), 31–55. https://doi.org/10.1177/0162353220978304

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Prast, E. J., Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. (2015). Readiness-based differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*(2), https://doi.org/10.14786/flr.v3i2.163

Rakap, S., & Kaczmarek, L. (2010). Teachers' attitudes towards inclusion in Turkey. *European Journal of Special Needs Education*, *25*(1), 59–75. https://doi.org/10.1080/08856250903450848

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*(11), 1186–1204. https://doi.org/10.1080/13603116.2012.743604

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171–195). Routledge.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321. https://doi.org/10.1177/0734282911406653

Streiner, D. L., Norman, G. R., & Cairney, J. (2024). *Health measurement scales: a practical guide to their development and use* (6th ed.). Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Tant, M., & Watelain, E. (2016). Forty years later, a systematic literature review on inclusion in physical education (1975-2015): A teacher perspective. *Educational Research Review*, *19*, 1–17. https://doi.org/10.1016/j.edurev.2016.04.002

Tarantino, G., Makopoulou, K., & Neville, R. D. (2022). Inclusion of children with special educational needs and disabilities in physical education: A systematic review and meta-analysis of teachers' attitudes. *Educational Research Review*, *37*, Article 100456. https://doi.org/10.1016/j.edurev.2022.100456

Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 414–424. https://doi.org/10.1037/0022-0167.34.4.414

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. ASCD.

Tomlinson, C. A., & Imbeau, M. B. (2023). *Leading and managing a differentiated classroom*. ASCD.

Turner, D. P. (2020). Sampling methods in research design. *Headache: The Journal of Head and Face Pain, 60*(1), 8–12. https://doi.org/10.1111/head.13707

UNESCO. (2019). *On the road to inclusion: Highlights from the UNICEF and IIEP technical round tables on disability-inclusive education sector planning*. UNESCO.

Van Geel, M., Keuning, T., & Safar, I. (2022). How teachers develop skills for implementing differentiated instruction: Helpful and hindering factors. *Teaching and Teacher Education: Leadership and Professional Development*, *1*, Article 100007. https://doi.org/10.1016/j.tatelp.2022.100007

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

Vita, S. C. (2023). The problematisation of ethnic and cultural diversity in physical education teacher education (PETE): An analysis of PETE course syllabi from Norway, Aotearoa/New Zealand and Canada. *Sport, Education and Society*, *28*(10), 1–15. https://doi.org/10.1080/13573322.2023.2284804

Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment, 16*(3), 231–243. https://doi.org/10.1037/1040-3590.16.3.231

Wen, Q., & Cai, J. (2024). Applying structural equation modeling to examine the role of teacher beliefs and practices in differentiated instruction in physical education: Multiple mediation analyses. *Psychology in the Schools*, *61*(7), 3045–3062. https://doi.org/10.1002/pits.23206

Whipp, P., Taggart, A., & Jackson, B. (2014). Differentiation in outcome-focused physical education: Pedagogical rhetoric and reality. *Physical Education and Sport Pedagogy*, *19*(4), 370–382. https://doi.org/10.1080/17408989.2012.754001

Wilhelmsen, T., Sørensen, M., & Seippel, Ø. N. (2019). Motivational pathways to social and pedagogical inclusion in physical education. *Adapted Physical Activity Quarterly*, *36*(1), 19–41. https://doi.org/10.1123/apaq.2018-0019

Wilkinson, S. D., & Penney, D. (2023). A national survey of gendered grouping practices in secondary school physical education in England. *Physical Education and Sport Pedagogy*, *28*(5), 1–16. https://doi.org/10.1080/17408989.2023.2236642

Wuensch, K. L. (2019). *Coefficient omega: Wuensch's R lessons*. http://core.ecu.edu/psyc/wuenschk/R-Lessons/Omega_McDonald.pdf

Yuen, S. Y., Leung, C. C. Y., & Wan, S. W.-Y. (2022). Teachers' perceptions and practices of differentiated instruction: Cross-cultural validation of the differentiated instruction questionnaire in Hong Kong. *International Journal of Educational Research*, *115*, Article 102044. https://doi.org/10.1016/j.ijer.2022.102044

Appendix A

Comparison of Existing Differentiated Instruction Measurement Instruments

Authors (Instuments)	Context	Language	Factors (items)	Samples	Key Findings
Roy et al. (2013) DIS	Primary education (French/ Math)	French	2-factor (12)	*n*= 125	Good structural validity; teachers prefer low-preparation practices
Prast et al. (2015) DSAQ	Primary mathematics	Dutch	2-factor (56)	*n*= 268	Better fit than 5-factor; moderate-strong validity
Van Geel et al. (2022) DSAQ+	Primary mathematics	Dutch	5-factor (56)	*n*= 288	Lower implementation frequency; greater perceived challenge
Coubergs et al. (2017) DI-QUEST	Primary/ Secondary	Dutch	5-factor (31)	*n*= 1,574	Flexible grouping predicts DI; curriculum adherence negative
Yuen et al. (2022) DI-QUEST (HK)	Primary/ Secondary	Chinese	5-factor (31)	*n* = 416	Confirmed structure; mindset highest, ethics lowest
Wen & Cai (2024) DI-QUEST (PE)	PE (preservice)	Chinese	5-factor (31)	*n* = 527	Growth mindset most important; output/input non-significant
Pereira et al. (2021) CPS-Revised	General education	English	4-factor (various)	*n* = 648	Weaker validity for low achievers; practice variability important
Suprayogi et al. (2017) DIIS	General education	English	4-factor (15)	-	High reliability (α = .92)
Lieberman et al. (2019) LIRSPE	Physical Education	English	Observation (rating scale)	PE	Face/content validity; test-retest reliability

Appendix B

Preliminary (n=37) Version of the Questionnaire and Phase 1 Clarifications following Cognitive Interviews

Directives :
Guidelines:
En considérant vos pratiques d'enseignement habituelles, indiquez :
Considering your usual teaching practices, indicate:

À quel point vous utilisez les stratégies suivantes (fréquence : 1= jamais, 5 = très souvent);
To what extent do you use the following strategies (frequency: 1 = never, 5 = very often);

À quel point vous jugez votre niveau de compétence pour chacune des pratiques (1= très bas, 5 = très élevé).
To what extent do you assess your level of competence for each practice (1 = very low, 5 = very high).

Item #	Original Dimensions and items	Phase 1 Clarifications following group and individual cognitive interviews
	Connaitre les besoins hétérogènes - **Considering the heterogeneity**	
1	Proposer des tâches en fonction des intérêts des élèves. **Determine tasks based on students' interests**.	
~~20~~	~~Proposer des tâches en fonction des choix des élèves.~~ ~~**Determine tasks based on students' choice.**~~	
4	Proposer des tâches en fonction du genre des élèves. **Determine tasks based on students' gender.**	
~~8~~	~~Proposer des tâches en fonction du niveau des habiletés des élèves.~~ ~~**Determine tasks based on students' readiness.**~~	
~~12~~	~~Proposer des tâches en fonction des besoins particuliers des élèves.~~ ~~**Determine tasks based on students' special educational needs.**~~	
13	Proposer des tâches en fonction des préférences des élèves. **Determine tasks based on students' preferences.**	
15	Proposer des tâches en fonction du milieu culturel des élèves. **Determine tasks based on students' cultural background.**	

	Faire des ajustements pédagogiques - **Adjusting practices to handle heterogeneity**	
2	Varier les modalités de réalisation des tâches pour tenir compte des différences des élèves. **Adjust the tasks' parameters to accommodate for the differences among students.**	(changer le nombre de joueurs, changer les règles ou l'aire du jeu, etc.) **(change the number of players, change the rules or the game area, etc.)**
~~3~~	~~Fournir des outils pédagogiques supplémentaires qui tiennent compte des différences et des besoins des élèves.~~ **~~Provide additional educational tools that consider the differences and needs of the students.~~**	~~(plan de travail, démonstrations, support visuel, etc.).~~ **~~(work plan, demonstrations, visual aids, etc.)~~**
~~11~~	~~Ajuster les tâches évaluatives afin de tenir compte des différences et des besoins des élèves.~~ **~~Adjust the assessment tasks to consider the differences and needs of the students.~~**	
~~5~~	~~Proposer du matériel varié et adapté en fonction des différences et des besoins des élèves.~~ **~~Provide varied and adapted materials based on the differences and needs of the students~~.**	
6	Ajuster la quantité de travail en fonction des capacités des élèves. **Adjust the workload to the students' skills.**	
~~7~~	~~Proposer des regroupements de travail flexibles et évolutifs pour répondre aux besoins des élèves.~~ **~~Propose flexible and adaptable work groupings to meet the needs of the students.~~**	~~(groupes homogènes, hétérogènes, dyade, travail seul)~~ **~~(homogeneous groups, heterogeneous groups, pairs, individual work)~~**
~~14~~	~~Planifier des dyades d'entraide ou du tutorat par les pairs.~~ **~~Plan dyads for peer support or peer tutoring.~~**	
16	Utiliser différentes méthodes pour présenter les contenus d'apprentissage **Use different methods of presentation.**	(démonstration, vidéo, schéma) **(demonstration, video, diagram)**
~~22~~	~~Expliquer une notion à un petit groupe d'élèves ayant ponctuellement besoin d'un soutien particulier pendant que les autres réalisent une tâche.~~ **~~Explain a concept to a small group of students who occasionally need support while the others complete a task.~~**	
23	Varier le degré de complexité ou d'intensité d'une tâche afin de tenir compte des besoins des élèves. **Adjust a task's difficulty or intensity to accommodate students' needs.**	
~~27~~	~~Planifier des tâches qui privilégient l'apprentissage coopératif.~~ **~~Plan tasks that prioritize cooperative learning.~~**	

28	Proposer différentes modalités de tâches évaluatives. **Varied modalities for assessment tasks.**	(vidéos, photos, discussion en privé, en groupe, individuel, présentation orale, par les pairs, auto-évaluation, etc.) **(videos, photos, private discussion, in groups, individually, oral presentations, peer evaluation, self-assessment, etc.)**
29	Proposer à l'ensemble du groupe, différents défis d'apprentissage afin que chaque élève puisse choisir celui qui lui convient. **Provide a range of learning challenges allowing student to select the one that best meets their needs.**	
~~26~~	~~Adapter les documents remis aux élèves selon leur niveau de compréhension de l'information.~~ **~~Adapt the documents provided to students according to their level of understanding of the information.~~**	~~(ex. regrouper l'information différemment, donner plus de détails)~~ **~~(e.g., grouping the information differently, providing more details)~~**
~~31~~	~~Offrir des choix aux élèves pour faire la démonstration de leurs compétences.~~ **~~Offer students choices to demonstrate their skills.~~**	
~~32~~	~~Aménager l'environnement d'apprentissage en organisant différents ateliers de travail.~~ **~~Arrange the learning environment by organizing different workstations.~~**	~~(coin calme, explications, visionnement, corrections, etc.).~~ **~~(quiet corner, explanations, viewing, corrections, etc.).~~**
33	Permettre aux élèves de recourir de manière autonome aux ressources et au matériel dont ils ont besoin pour réaliser la tâche. **Facilitate students' autonomous access to the resources and materials for assignment completion.**	

Suivre les progrès des élèves~~ tous et~~ élèves ayant des besoins particuliers (EBP)- **Monitoring the progress of students having special educational needs (SEN)**

~~9~~	~~Utiliser les données issues de l'évaluation diagnostique afin d'ajuster les tâches d'apprentissage pour l'ensemble du groupe.~~ **~~Use the data from the diagnostic assessment to adjust the learning tasks for the entire group.~~**	
~~17~~	~~Réguler fréquemment le progrès de l'ensemble du groupe d'élèves.~~ **~~Frequently regulate the progress of the entire group of students.~~**	~~(aide à l'apprentissage)~~ **~~(learning support)~~**
~~19~~	~~Consigner les ajustements ou des adaptations proposés aux élèves.~~ **~~Record the adjustments or adaptations proposed to the students.~~**	

~~21~~	~~Utiliser des outils ou moyens d'observation variés afin d'identifier les différences, les besoins ou le niveau de compétences des élèves.~~ **~~Use various tools or observation methods to identify the differences, needs, or skill levels of the students.~~**	
24	Utiliser mes observations sur les progrès du groupe pour prendre des décisions quant aux ajustements à apporter à l'évaluation. **Adjust the assessment task based on the observations of the group's progress.**	
~~34~~	~~Vérifier si les interventions auprès de l'ensemble du groupe sont efficaces.~~ **~~Check if the interventions with the entire group are effective.~~**	~~(p. ex. contrôler le progrès du groupe à la suite d'un ajustement des tâches).~~ **~~(e.g., monitor the group's progress following an adjustment in tasks).~~**
~~36~~	~~Faire appel à des ressources externes pour vous aider à connaître les besoins de l'ensemble des élèves du groupe.~~ **~~Call upon external resources to help you understand the needs of all the students in the group.~~**	~~(conseillers pédagogiques, titulaires, collègues, experts externes, direction, éducateurs spécialisés, psychoéducateurs, etc.)~~ **~~(educational advisors, homeroom teachers, colleagues, external experts, administration, specialized educators, psychoeducators, etc.).~~**
~~10~~	~~Utiliser les données issues de l'évaluation diagnostique afin d'ajuster les tâches d'apprentissage spécifiquement pour les ÉBP.~~ **~~Use the data from the diagnostic assessment to specifically adjust the learning tasks for the EBP.~~**	
18	Réguler fréquemment le progrès spécifiquement des ÉBP. **Monitor the progress of students who require special learning support.**	(aide à l'apprentissage) **(learning support)**
25	Utiliser mes observations sur les progrès d'un ÉBP pour prendre des décisions quant aux ajustements à apporter à son évaluation. **Adjust one student's assessment task based on his progress**	
30	Proposer des défis d'apprentissage différents spécifiquement pour les ÉBP. **Propose distinct learning tasks for students having special educational needs.**	
35	Vérifier spécifiquement si les interventions auprès des ÉBP sont efficaces. **Evaluate the efficacy of the interventions provided to students with special educational needs**.	(ex. évaluer le progrès des ÉBP réalisé à la suite d'un ajustement de l'enseignement ou d'une tâche).
~~37~~	~~Faire appel à des ressources externes pour vous aider à connaître spécifiquement les besoins des ÉBP.~~ **~~Call on external resources to help you specifically understand the needs of SEN.~~**	~~(conseillers pédagogiques, titulaires, collègues, experts externes, direction, éducateurs spécialisés, psychoéducateurs, etc.).~~

	~~(educational advisors, homeroom teachers, colleagues, external experts, administration, specialized educators, psychoeducators, etc.).~~

Note. The ~~crossed-out~~ statements were not retained following the validity analysis. This is a free translation by the authors. English items have not been validated