Measuring student engagement in lessons using an experience sampling methodology : the development and validation of the dynamic engagement with learning questionnaire

Reference:
De Weerdt Dries, Simons Mathea, Struyf Elke.- Measuring student engagement in lessons using an experience sampling methodology : the development and validation of the dynamic engagement with learning questionnaire
Journal of psychoeducational assessment - ISSN 0734-2829 - (2024), p. 1-13
Full text (Publisher's DOI): https://doi.org/10.1177/07342829241240901
To cite this reference: https://hdl.handle.net/10067/2050440151162165141

Measuring Student Engagement in Lessons Using an Experience Sampling Methodology: The Development and Validation of the Dynamic Engagement with Learning Questionnaire

Dries De Weerdt, Mathea Simons, and Elke Struyf

Antwerp School of Education, Faculty of Social Sciences, University of Antwerp (Belgium)

Author note

Dries De Weerdt https://orcid.org/0000-0001-6903-6587

Mathea Simons https://orcid.org/0000-0002-9239-9324

Elke Struyf https://orcid.org/0000-0003-1067-6357

We have no known conflict of interest to disclose. We received funding from Fonds Wetenschappelijk Onderzoek (FWO), the Research Foundation Flanders, as this study is part of the larger ESTAFETT project (www.teamteaching-estafett.be/english).

Correspondence concerning this article should be addressed to Dries De Weerdt, University of Antwerp, Sint-Jacobstraat 2, 2000 Antwerp, Belgium. Email: dries.deweerdt@uantwerpen.be

Abstract

Student engagement is a central construct in education research, as it is a strong predictor and/or mediator of both cognitive and non-cognitive student outcomes. However, there remain questions among researchers about how to properly measure student engagement during lessons. Our study contributes to this discussion by presenting the development and validation of a theoretically grounded questionnaire of students' lesson engagement, which is based on an experience sampling methodology. To validate this questionnaire, we issued it to 410 fifth- and sixth-grade students (10–12 years old). The results of both an exploratory and multilevel confirmatory factor analysis approved construct validity of the measure. Adequate within-person variability across time points was also found, which makes the questionnaire suitable for capturing fluctuations over time. Furthermore, partial scalar measurement invariance for questionnaire administration across time points and between fifth- and sixth-grade students was found.

Keywords: dynamic engagement with learning, students' lesson engagement, experience sampling, lesson practices, questionnaire development and validation

Measuring Student Engagement in Lessons Using an Experience Sampling Methodology: The Development and Validation of the Dynamic Engagement with Learning Questionnaire

Student engagement during lessons is critical for academic success and constitutes an essential precondition for optimal learning (Lei et al., 2018; Skinner et al., 2008). In general, students' lesson engagement describes the degree or quality of students' involvement in an activity or the learning environment—how productively they participate in learning opportunities during their lessons (Hofkens & Ruzek, 2019; Skinner et al., 2009).

A variety of indicators, instruments, and methods has been applied to capture students' lesson engagement (Sinatra et al., 2015). All existing theoretical and methodological approaches consider engagement as a multidimensional construct with several underlying indicators (Fredricks et al., 2004; Hofkens & Ruzek, 2019).

From a conceptual point of view, one of the most prominent frameworks to capture engagement stems from the work of Skinner et al. (2008), who have applied a motivational conceptualization of engagement in the classroom. They have recognized the complexity of lesson engagement, taking its multifaceted character into account and distinguishing between two dimensions: behavioral and emotional engagement. Furthermore, they have acknowledged the negative counterpart of engagement: disaffection. Disaffection includes both the absence of positive engagement and the presence of negative engagement (Skinner et al., 2009). Table 1 presents a short definition of each of these components.

Table 1
Conceptualization of Student Engagement during Lessons according to Skinner et al. (2008)

Component	Definition	Example indicators
Behavioral engagement	Students' explicit participation in learning activities during a lesson.	effort, attention, persistence
Behavioral disaffection	Students' passivity in learning activities during lessons.	distraction, withdrawal, passivity

Emotional engagement	Students' appreciation and affective responses toward learning activities during a lesson.	pleasure, enthusiasm, interest
Emotional disaffection	Students' negative emotions toward learning activities during lessons.	boredom, anxiety, frustration

Methodologically, several approaches exist for understanding student engagement during lessons. For example, teacher questionnaires have been developed to assess students' engagement (e.g., Hart et al., 2011). An advantage of this type of measure is that information on engagement can be obtained regardless of student characteristics (e.g., language deficiencies). However, it may be difficult to report on students' emotional engagement, as it requires teachers to indicate how students feel during lessons (Wang et al., 2016). It can also be challenging for teachers to distinguish different levels of engagement for all students in a class.

Another way to measure lesson engagement is to apply interview methods with students (Fredricks et al., 2016). The advantage of this method is that the interviewer can provide additional explanations and clarify ambiguities (for example when students don't understand a question). Also, students can describe their engagement and their experiences in their own words.

Further, students' lesson engagement can also be observed. Observational measures have the benefit of assessing engagement as it occurs in context, which provides opportunities to unravel the association between engagement and specific lesson events (Pianta, 2016). On the other hand, observational methods are time-intensive and its results may be more difficult to generalize (Fredricks & McColskey, 2012).

However, the most common way to measure engagement is by means of student self-report questionnaires (see Fredricks & McColskey, 2012, for an overview of possible measurement instruments). Student questionnaires are convenient, easy to administer, and can be conducted in a variety of settings at a relatively low cost. Further, student reports have good face validity as students can provide reliable information into their own involvement during lessons (Appleton et al., 2008).

Engagement versus Disaffection with Learning (EDL) Questionnaire

In current educational literature, most questionnaires focus on engagement at the school level (see Fredricks & McColskey, 2012), which mostly ignores the specific class context (Hofkens & Ruzek, 2019). However, students' engagement during lessons is a key contributor to overall school success. Lesson engagement predicts the quality of learning, results on achievement tests, and academic resilience (Skinner et al., 2008).

One existing tool that aims to capture students' lesson engagement is the *Engagement versus Disaffection with Learning* questionnaire (EDL). Based on the four subcomponents (i.e., behavioral engagement, behavioral disaffection, emotional engagement, and emotional disaffection), Skinner et al. (2008) developed the EDL, which, over the years, has been applied in education research worldwide in various situations with different groups of respondents (e.g., Immekus & Ingle, 2019; Ritoša et al., 2020; Rodríguez-Medellín et al., 2020). The original EDL included 20 self-report items that used a 5-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree). Some example items are "I try hard to do well in class" (behavioral engagement), "When I'm in class, I think about other things" (behavioral disaffection), "Class is fun" (emotional engagement), and "When we work on something in class, I feel bored" (emotional disaffection). A four-factor structure with strong internal consistency (Cronbach's α > .70) among items was found (Skinner et al., 2008).

Although the EDL is considered a conceptually strong instrument, several authors in the field of education research argue that capturing students' lesson engagement convincingly remains a challenge (e.g., Appleton et al., 2006; Eccles, 2016). For example, one issue with traditional measuring methods, such as the EDL, is that they do not take intra-lesson fluctuations into account by only conducting questionnaires after a lesson has finished.

However, as many educational practitioners will recognize, students' engagement over the course of a lesson can vary considerably (Pöysä et al., 2018; Sulis, 2022; van Uden et al., 2013). Ignoring these possible fluctuations neglects the dynamic nature of engagement and its embeddedness in the ever-changing social context. Problems with recall bias may also arise, as students may not remember a past lesson event or experience accurately when they have to report about it (te Braak et al., 2023).

Experience Sampling Method (ESM)

To overcome these challenges, methods are developed to capture the dynamics and fluctuations of engagement over shorter periods of time. Unlike other measurement methods, dynamic measures of engagement provide rich information about student engagement in context, particularly how engagement fluctuates during or across tasks, social interactions, and learning environments (Hofkens & Ruzek, 2019).

A recommended approach for assessing the dynamic nature of engagement is to make use of the experience sampling method (ESM), which refers to collecting real-time information during lessons (Bolger & Laurenceau, 2013; te Braak et al., 2023). Compared to traditional post-lesson questionnaire measures, using this method to research engagement during lessons can provide more valid and reliable information about specific activities and specific contexts. In addition, student self-reports of engagement gathered using the ESM are more reliable as they are collected in, or close to, the moment of interest, reducing issues with recall (Sinatra et al., 2015). Typical self-report questionnaires conducted at the end of a lesson cannot effectively measure the influence of contextual features that impact engagement in the moment of learning. ESM provides a person-in-context approach of measuring engagement, which gives evidence for strong ecological validity of the questionnaire due to the increase of students' sensitivity to the questionnaire items as they can better reflect on their momentary feelings (Larson & Csikszentmihalyi, 2014; Sinatra et al., 2015). Collecting ESM data also creates opportunities to compare the results of particular lesson phases and activities. This method can thereby advance our understanding of the momentary shifts in student engagement during and across lessons.

The Present Study

Despite the potential advantage of the ESM for capturing students' lesson engagement data (Eccles, 2016; Sinatra et al., 2015), few psychometric studies of the measures utilized in ESM research have been conducted. In addition, existing tools, such as the EDL, may not be appropriately reliable and valid measures for ESM studies, as they are often not designed to repeatedly capture momentary experiences (Eisele, Vachon, et al., 2022). This means that there would be no guarantee that they would function as intended in experience sampling studies that seek to capture fluctuating processes. Therefore, a comprehensive validation process of new instruments to measure engagement with ESM is highly recommended.

To fill this gap, our study aims to further conceptualize, test, and validate scores from an instrument designed to measure student engagement during lessons. We first adapted the original EDL so it can be efficiently used with the ESM while maintaining its conceptual credibility. Second, we validated the questionnaire by using it with fifth- and sixth-grade students (10–12 years old).

Method

The questionnaire development and validation process involved three phases, each with a specific purpose. Before the first phase began, the entire study was approved by the ethical advisory committee at the authors' university. All of the teachers, students, and parents gave their active consent to participate in the study.

Phase 1: Development

The starting point of developing a new ESM measure for engagement was the Dutch version of the EDL, which has been applied in many previous studies (e.g., Vandenkerckhove et al., 2019). As previously mentioned, the original EDL consisted of 20 self-report items (Skinner et al., 2008). However, we considered a 20-item questionnaire to be too long to conduct multiple times during a lesson without disturbing the lesson dramatically. To reduce the number of items, we used the general principles outlined by Eisele, Kasanova, et al. (2022) for ESM questionnaire development.

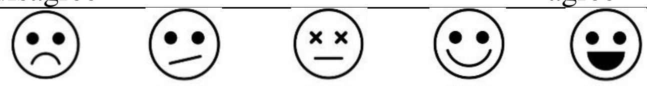
First, we selected items that capture dynamic phenomena that can change quickly (i.e., during a lesson). As we aimed to identify fluctuations in engagement over a single lesson, relatively stable indicators (e.g., a student's interest in the topic) were inconvenient and not meaningful for this research. On the contrary, indicators such as pleasure and distraction were highly suitable, as they are specific to a present moment with potential within-person variability over a short period. For these reasons, we choose four indicators of engagement, one for each of the subcomponents of engagement (see Table 1): (1) effort, (2) distraction, (3) pleasure, and (4) boredom. All of these indicators have a strong dynamic character. The factor loadings of previous research have also shown strong connections with their respective underlying subcomponents (e.g., Immekus & Ingle, 2019; Rodríguez-Medellín et al., 2020).

Second, when formulating ESM items, they should include a reference to a specific time. We chose to formulate our items by referring to a specific time interval (e.g., "During the previous phase of the lesson, I was bored"). Thereby, we aimed to evaluate engagement after a specific (and short) lesson phase, allowing us to capture the entire lesson after administrating the questionnaire several times (four times per 60-minute lesson). In the setting of our study, these phases were easily distinguishable and involved four stages: (1) interactive introduction to the lesson theme, (2) classroom instruction, (3) station teaching, and (4) interactive discussion. When using the questionnaire in a different teaching context, the phases should be determined based on content being taught or activities being deployed. The lesson structure determines how many phases there are in a lesson and how often the questionnaire can be administered. Therefore, the number of times the questionnaire is administered during a lesson is not a fixed number.

Third, the wording of the items and the response scale options are crucial. Participants should be able to read and answer the chosen questions and response options quickly. We also avoided confusing wording because our target population included primary school children. The new questionnaire is presented in Table 2. We called it the *Dynamic Engagement with Learning* questionnaire (DEL).

Table 2

Dynamic Engagement with Learning questionnaire (DEL)

Item	Indicator	Response options				
The previous lesson phase…		Totally disagree	Disagree	Neutral	Agree	Totally agree
1. I tried hard to do well.	Behavior engagement					
2. I was thinking about other things.	Behavioral disaffection					
3. was fun.	Emotional engagement					
4. I felt bored.	Emotional disaffection					

Phase 2: Pre-Testing Study

During the pre-testing study, the DEL was administered to 46 primary school students in two classes in February 2023. The study design encompassed an event-contingent ESM design, wherein participants had to complete a questionnaire every time a pre-defined event (e.g., a lesson phase) was finished (Bolger & Laurenceau, 2013). At the beginning of the lesson, students received brief instructions (approximately one minute) on how to properly respond to the items (e.g., only one answer per item, answer honestly). During the lesson, students had a small workbook, with four copies of the questionnaire on a single page. After each lesson phase (e.g., interactive classroom instruction and discussion, station teaching), which lasted approximately 15 minutes, students completed the four items on engagement. In total, the questionnaire was administered four times during one 60-minute lesson. On average, students took approximately 30 seconds to complete the four items.

During this pre-testing phase, the feasibility of completing the questionnaire during lessons was evaluated. In this regard, after the lesson, a small group discussion with the whole class was conducted to retrieve information on ambiguous wording, the clarity of the response categories, and inclusion of important and fluctuating feelings. In general, the questionnaire was found to be highly effective and pleasant; no noteworthy problems were identified.

Phase 3: Validation Study

Finally, the DEL was administered in May 2023 to 20 classes at five schools, all identified using a purposive sampling approach. The first school served as an internal pilot study (*n* = 97), whereas the remaining four schools constituted the sample of the main study (*n* = 267). To maintain comparability, all data were collected during a lesson in the natural sciences. Table 3 presents additional information about the sample of the validation study and each of the four time points. In general, there was a fairly equal distribution of fifth- (10–11 years old) and sixth-grade (11–12 years old) students. Furthermore, the response rate of the participating students was high, with the greatest missingness (10%) at the final time point (T4).

Table 3
Study Sample Information

Study	Total Sample	Grade		Sample per time point (dropout rate)			
		5th grade	6th grade	T1	T2	T3	T4
Pilot study	97	52 (54%)	45 (46%)	96 (1%)	95 (2%)	97 (0%)	92 (5%)
Main study	267	139 (52%)	128 (48%)	254 (5%)	251 (6%)	255 (4%)	240 (10%)

To validate the questionnaire statistically, we applied both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The EFA was conducted using the sample from the internal pilot study to establish an initial factor model. Subsequently, the CFA assessed whether the proposed initial model was plausible using data from the main study (Lorenzo-Seva, 2022).

Exploratory Factor Analysis (EFA)

To determine the number of factors underlying the data, several procedures and criteria were used as part of the EFA process. First, the rule that eigenvalues must be greater than one was applied (Kaiser, 1960). Second, parallel analysis was used to compare the eigenvalues of the original data set and those of random simulated data. Factors that had greater eigenvalues when using the original data than those of the simulated data were considered for retainment (Preacher & MacCallum, 2003). Third, to evaluate the items of the factor solution, the approximate simple structure was considered. This implies that items were retained when they loaded strongly (loading > .40) on one factor (McDonald, 1985). The EFA allowed factors to be intercorrelated (i.e., oblimin rotation).

Multilevel Confirmatory Factor Analysis (ML-CFA)

To disentangle the between-person and within-person factor structure, a multilevel approach to CFA was used. The ML-CFA was tested using robust maximum likelihood estimation (MLR; Rhemtulla et al., 2012). MLR can handle missing data, which is often an issue in data collection methods with repeated measures, such as the ESM.

To test model fit, two incremental fit indices and two absolute model fit indices were examined: (1) Comparative Fit Index (CFI), (2) Tucker-Lewis Index (TLI), (3) Root Mean-Square Error of Approximation (RMSEA), and (4) Standardized Root Mean-Square Residual (SRMR). The cutoff criteria for an adequate model fit were a CFI and TLI greater than or equal to .90, and RMSEA and a SRMR lower than or equal to .08 (Whittaker & Schumacker, 2022). If a model showed a poor fit with the data, the modification index was used to identify suggestions for model improvement based on parameters that decrease the Chi-squared value (Marcoulides & Drezner, 2003). A suggested parameter was only included in the model if it was theoretically justifiable. To ensure that an increase in model complexity was better, two alternative models were compared using the Akaike's Information Criterion (AIC) and the Chi-square value (χ^2). The model with the lowest AIC and Chi-square value was considered best (Dziak et al., 2019).

In addition, we calculated the intraclass correlation (ICC) to check whether the items changed within persons enough to justify assessing them multiple times in a single lesson. The ICC measures how much of the variance is due to between-person differences compared to within-person fluctuations. It is calculated by dividing the between-person variance by the total variance. For example, an ICC of 0.4 indicates that 40% of the variance of that item is caused by between-person differences and 60% is caused by momentary fluctuations within students (Eisele, Kasanova, et al., 2022). A high ICC indicates low within-person variability and leads to reservations about measuring the variable multiple times in a short period of time (Bolger & Laurenceau, 2013).

For estimating reliability, both Cronbach's alpha and McDonald's omega estimates were used (Hayes & Coutts, 2020). Alpha (α) and omega (ω) values equal to or greater than .70 were considered acceptable.

Measurement Invariance

As we collected data from fifth- and sixth-grade students at different time points, we aimed to examine if our measurements were invariant for these variables, which would provide additional support for construct validity (Whittaker & Schumacker, 2022). A multiple group factor analysis was conducted, following a sequential process. In this process, models with increasingly restrictive equality constraints for model parameters across groups were consecutively checked: Model_1 checked whether the same factor structure held across all groups (i.e., configural invariance); Model_2 added the requirement of equal factor loadings across groups (i.e., metric or weak invariance); Model_3 included an additional constraint for equal intercepts across groups (i.e., scalar or strong invariance). These models were compared using the Chi-square difference test ($\Delta\chi^2$). A significant Chi-square test suggests the non-invariance of one or more parameters across groups. If non-invariance was found, a partial measurement invariance model was tested (Model_4), which released some of the constrained parameter estimates that functioned differently between groups. All models were estimated in R software with the aid of the "Lavaan" (Rosseel, 2012) and "Psych" (Revelle, 2023) packages.

Results

Descriptive Statistics

Table 4 presents item means and standard deviations for each time point, along with the ICCs and reliability estimates. The means and standard deviations of the engagement items suggest that students were highly engaged at all time points of the lesson. We found high values for the indicators "Effort" and "Pleasure" (M = 3.89–4.41) and lower values for "Distraction" and "Boredom" (M = 2.01-2.62). In addition, the standard deviations indicate that there was more variation between students at later time points.

Table 4
Descriptives

Item	Mean (standard deviation)				ICC	α if item dropped*
	T1	T2	T3	T4		
Effort	4.41 (0.69)	4.18 (0.77)	4.12 (0.81)	4.39 (0.76)	.38	.75
Distraction	2.62 (1.07)	2.25 (1.13)	2.30 (1.15)	2.16 (1.21)	.47	.72
Pleasure	3.89 (0.88)	4.12 (0.93)	3.98 (1.00)	4.10 (1.03)	.37	.66
Boredom	2.25 (1.01)	2.01 (1.06)	2.11 (1.14)	1.98 (1.12)	.44	.61

Note. Scale ranging for min. 1 to max. 5; *General estimate of reliability: α = .75, ω = .81

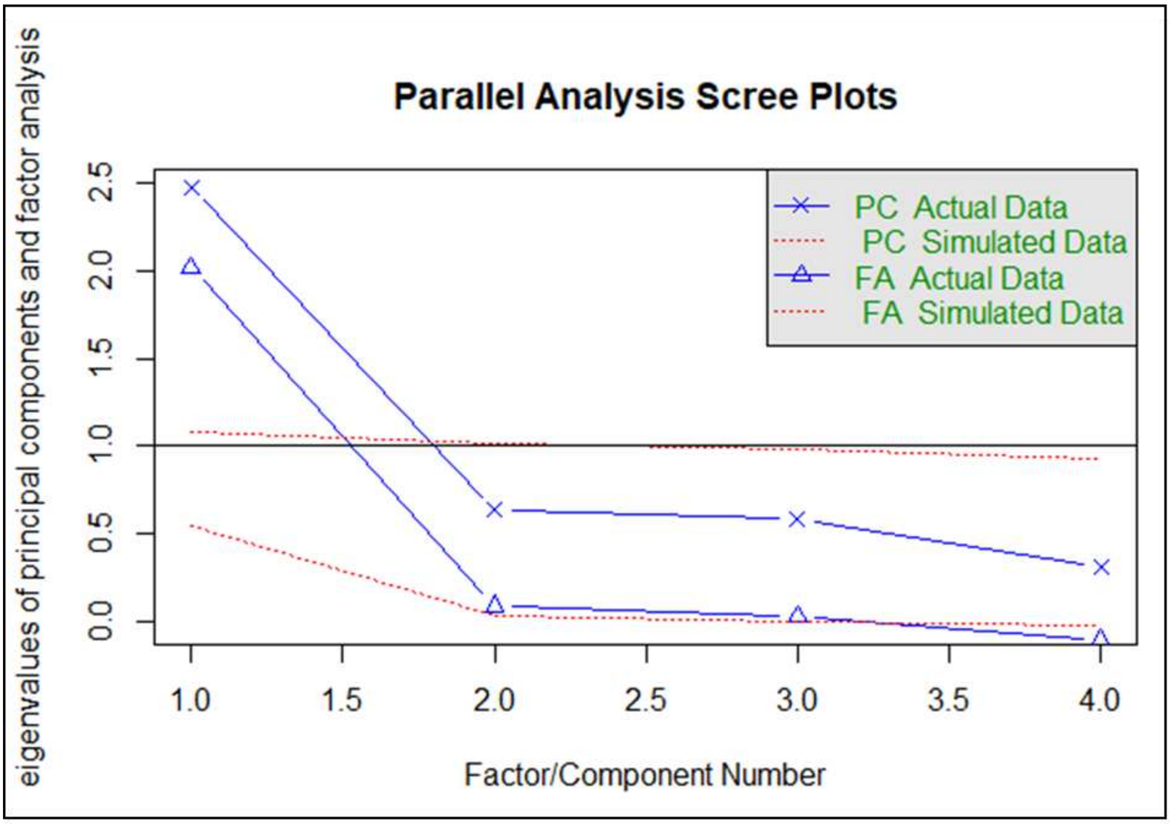
The ICCs suggested that most of the variation occurred within rather than between students. For example, the ICC for the indicator "Effort" was .38, which means that 38% of the variance was due to between-person variance, and the remaining 62% was within-person variance.

For reliability, the alpha and omega estimates were .75 and .81, respectively, which indicate an acceptable level of internal consistency for the items. Removing one of the items would not have resulted in a better alpha estimate.

EFA

The Kaiser criterion indicated that engagement has a one-factor structure, as only one eigenvalue was greater than one. On the contrary, the parallel analysis suggested that the first three eigenvalues for the sample correlation matrix were larger than those of the simulated data (see Figure 1). However, the three-factor structure was not desirable because Item 1 (Effort) did not load sufficiently on any of the factors (loading < .40). The one-factor structure was found to be appropriate, with relatively strong factor loadings, ranging from .57 (Effort) to .89 (Boredom) on the underlying factor (i.e., Engagement).

Figure 1
Scree Plot



Multilevel CFA

Building on the EFA results, an ML-CFA was specified for engagement as the overall latent factor. For the initial model, some of the model fit indices showed an inadequate fit with the data (CFI = .96, TLI = .88, RMSEA = .09, $SRMR_{within}$ = .02, $SRMR_{between}$ = .08). Therefore, we retrieved the modification index and followed its recommendation to include the error-covariance between the indicators “Effort” and “Pleasure” and thus improve the model fit. All of the fit indices for the modified model indicated a proper fit (CFI = .99, TLI = .96, RMSEA = .05, $SRMR_{within}$ = .02, $SRMR_{between}$ = .07). Comparing the initial model and the modified model, the AIC and Chi-square were lower for the modified model (ΔAIC = 44, $\Delta\chi^2$ = 99.65), indicating that it was more suitable.

Table 5 presents all of the parameter estimates at the measurement level. The R-square values indicate the proportion of variance in the indicator variable explained by the model. In our data, the most variance (89%) was explained for the indicator ‘Boredom” (at the between level) and the least (17%) was explained for the indicator “Effort” (at the within level).

Table 5

Parameter Estimates for the Measurement Model

Item	Unstandardized factor loadings	S.E.	Standardized factor loadings	R^2
Within level				
1. Effort	1.00		.40	.17
2. Distraction	-1.60	.15	-.46	.21
3. Pleasure	2.17	.14	.69	.47
4. Boredom	-2.59	.22	-.77	.59
Between level				
1. Effort	1.00		.64	.41
2. Distraction	-2.05	.36	-.77	.60
3. Pleasure	1.50	.15	.76	.58
4. Boredom	-2.31	.40	-.94	.89

Note. All parameter estimates were statistically significant with $p < .05$

Measurement Invariance

Table 6 presents the sequence of model fits used for testing measurement invariance between (1) the four time points during the lesson and (2) the two grade levels. Examining the fit indices, most of the tested models show an acceptable model–data fit. Only the RMSEA values were above the required .08 value. For both time points and grade, Model_3, which tested for scalar invariance, was not invariant based on the significant Chi-square test. However, partial scalar invariance was found when we released the constraints for the indicator "Effort" (Model_4).

Table 6

Measurement Invariance for Time Point and Grade

	χ^2	df	AIC	BIC	CFI	TLI	RMSEA	SRMR
Time point								
Model 1 (Configural)	118.55	8	20,532	20,802	.95	.84	.16	.04
Model 2 (Metric)	129.98	17	20,526	20,745	.95	.93	.11	.04
Model 3 (Scalar)	251.10*	26	20,629	20,797	.90	.90	.13	.07
Model 4 (Partial)	131.84	20	20,522	20,724	.95	.94	.10	.04
Grade								
Model 1 (Configural)	113.11	4	20,633	20,768	.95	.85	.15	.04
Model 2 (Metric)	120.88	7	20,634	20,753	.95	.91	.13	.04
Model 3 (Scalar)	132.27*	10	20,640	20,741	.94	.93	.11	.04
Model 4 (Partial)	122.13	9	20,632	20,739	.95	.93	.11	.04

Note. *$\Delta\chi^2$ test was statistically significant with $p < .05$

Discussion

The purpose of this study was to develop and validate a tool to measure student engagement during lessons, specifically for use in experience sampling research. The process involved construct definition, item generation and adaption, piloting, and factor analyses. The final result was the *Dynamic Engagement with Learning (DEL)* questionnaire, which is a 4-item instrument with a one-factor structure to capture student engagement during lessons. Student engagement was theoretically grounded in a motivational conceptualization of engagement and disaffection (Skinner et al., 2008). This conceptualization encompasses active involvement during lesson activities and the absence of behaviors and emotions that reflect a counterproductive motivational state.

The validation process showed a strong factor structure and adequate within-person variability. These properties are essential for a measure of lesson engagement used in experience sampling studies, which seek to understand momentary influences and fluctuations across time. Furthermore, partial strong invariance was found for different time points and grade levels (i.e., fifth- and sixth-grade students).

The use of our measure of students' lesson engagement can be of value for researchers who seek to examine (the evolution in) students' involvement during lessons. Also, our measure on engagement may be of added value for teachers and policymakers as it provides insight into which students make optimal use of lesson activities and when students are likely to lose involvement during the lesson (Hofkens & Ruzek, 2019). This information can contribute to evaluating the effectiveness of new educational interventions or teaching methods and determine which students need extra attention or support.

Limitations and Directions for Future Research

To further validate the use of the DEL questionnaire, studies with a different target population (e.g., other educational levels; different countries) can be conducted. The results of our study were obtained with fifth- and sixth-grade primary education students in Flanders (Belgium), which limits their generalizability to other populations or contexts.

Future research can also use different indicators for operationalizing engagement. This would create opportunities to compare the indicators in our study with different indicators used in other studies. However, as stated in the methodology, we chose the four indicators that were conceptually and methodologically the most convenient for measuring student engagement using the experience sampling method.

In addition, our study was limited to self-report measures from students. A major weakness of student self-report questionnaires during lessons is the likelihood of socially desirable answers. Therefore, in ideal circumstances, researchers triangulate information obtained from multiple measures of engagement (Fredricks & McColskey, 2012). In this regard, an opportunity for future studies is to translate the student version of the questionnaire to a teacher version or to combine the student questionnaire data with observation data. Combining data from different perspectives may enhance the amount of theoretical knowledge that can be retrieved from a certain study by not only relying on student self-report measures. Furthermore, a multimethod data collection can add evidence for convergent and discriminant validity of the instrument.

Further, a possible next step in studying students' lesson engagement using the DEL questionnaire is by looking at facilitators that may affect students' level of engagement. Our study was focused on the development and validation of the DEL. Thus, applying the questionnaire in studies seeking to understand the association between engagement and, for example, instructional strategies or student characteristics, would be valuable for both the research field on student engagement (e.g., gaining more fine-grained insight into the fluctuations of engagement during lessons) and for claiming reliability and validity of the instrument.

In conclusion, our study adds to the current knowledge base on educational measurement, as there is a need for psychometric studies on measures of student engagement during lesson practices using an experience sampling methodology. The findings of our study illustrate that our instrument can be used effectively to capture student engagement during lessons, and examine group differences and differences over time (Beaujean, 2014). By these means, researchers and educational practitioners can identify engagement, and its potential fluctuations, during the lessons in which they are interested.

References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5), 369–386. https://doi.org/10.1002/pits.20303

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2019). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, *43*, 71–75. https://doi.org/10.1016/j.learninstruc.2016.02.003

Eisele, G., Kasanova, Z., & Houben, M. (2022). Questionnaire design and evaluation. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed.). The Center for Research on Experience Sampling and Ambulatory Methods Leuven (REAL).

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and instruction*, 43, 1-4. https://doi.org/10.1016/j.learninstruc.2016.02.002

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_37

Hart, S. R., Stewart, K., & Jimerson, S. R. (2011). The student engagement in schools questionnaire (SESQ) and the teacher engagement report form-new (TERF-N): Examining the preliminary evidence. *Contemporary School Psychology*, 15(1), 67-79. https://doi.org/10.1007/BF03340964

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hofkens, T. L., & Ruzek, E. (2019). Measuring student engagement to inform effective interventions in schools. In J. A. Fredricks, A. L. Reschly, & S. L. Christenson (Eds.), *Handbook of Student Engagement Interventions* (pp. 309–324). Academic Press. https://doi.org/10.1016/B978-0-12-813413-9.00021-8

Immekus, J. C., & Ingle, W. K. (2019). A test and cross-validation of the factor structure of the engagement versus disaffection with learning instrument among middle school students. *Journal of Psychoeducational Assessment*, *37*(4), 504–511. https://doi.org/10.1177/0734282917742569

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Larson, R., Csikszentmihalyi, M. (2014). The Experience Sampling Method. In: *Flow and the Foundations of Positive Psychology.* Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9088-8_2

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, *46*, 517–528. https://doi.org/10.2224/sbp.7054

Marcoulides, G. A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 154–164. https://doi.org/10.1207/S15328007SEM1001_8

McDonald, R. P. (1985). Comments on D. J. Bartholomew, Foundations of factor analysis: Some practical implications. *British Journal of Mathematical and Statistical Psychology*, *38*(2), 134–137. https://doi.org/10.1111/j.2044-8317.1985.tb00827.x

Pianta, R. C. (2016). Teacher–Student Interactions: Measurement, Impacts, Improvement, and Policy. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 98-105. https://doi.org/10.1177/2372732215622457

Pöysä, S., Vasalampi, K., Muotka, J., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2018). Variation in situation-specific engagement among lower secondary school students. *Learning and Instruction*, *53*, 64–73. https://doi.org/10.1016/j.learninstruc.2017.07.007

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

Revelle, W. (2023). *Psych: Procedures for psychological, psychometric, and personality Research*. Northwestern University, Evanston, Illinois. R package version 2.3.3, https://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Ritoša, A., Danielsson, H., Sjöman, M., Almqvist, L., & Granlund, M. (2020). Assessing school engagement – adaptation and validation of "engagement versus disaffection with learning: Teacher report" in the Swedish educational context [Original Research]. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.521972

Rodríguez-Medellín, R., Zamarripa, J., Marentes-Castillo, M., Otero-Saborido, F., Baños, R., & Morquecho-Sánchez, R. (2020). Mexican validation of the engagement and disaffection in physical education scale. *International Journal of Environmental Research and Public Health*, *17*(6). https://doi.org/10.3390/ijerph17061821

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, *50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765–781. https://doi.org/10.1037/a0012840

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*(3), 493–525. https://doi.org/10.1177/0013164408323233

Sulis, G. (2022). Engagement in the foreign language classroom: Micro and macro perspectives. *System*, *110*, 102902. https://doi.org/10.1016/j.system.2022.102902

te Braak, P., van Tienoven, T. P., Minnen, J., & Glorieux, I. (2023). Data quality and recall bias in time-diary research: The effects of prolonged recall periods in self-administered online time-use surveys. *Sociological Methodology*, *53*(1), 115–138. https://doi.org/10.1177/00811750221126499

van Uden, J. M., Ritzen, H., & Pieters, J. M. (2013). I think I can engage my students. Teachers' perceptions of student engagement and their beliefs about being a teacher. *Teaching and Teacher Education*, *32*, 43–54. https://doi.org/https://doi.org/10.1016/j.tate.2013.01.004

Vandenkerckhove, B., Soenens, B., Van der Kaap-Deeder, J., Brenning, K., Luyten, P., & Vansteenkiste, M. (2019). The role of weekly need-based experiences and self-criticism in predicting weekly academic (mal)adjustment. *Learning and Individual Differences*, *69*, 69–83. https://doi.org/10.1016/j.lindif.2018.11.009

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction*, 43, 16-26. http://doi.org/10.1016/j.learninstruc.2016.01.008

Whittaker, T., & Schumacker, R. (2022). *A beginner's guide to structural equation modeling* (5th ed.). Routledge.