Testing the Validity and Reliability of Intrinsic Motivation Inventory Subscales within ASSISTments

Korinn S. Ostrow[1[0000-0001-7149-1802]] & Neil T. Heffernan[1[0000-0002-3280-288X]]

[1] Worcester Polytechnic Institute, Worcester MA 01609, USA
[ksostrow, nth]@wpi.edu

Abstract. Online learning environments allow for the implementation of psychometric scales on diverse samples of students participating in authentic learning tasks. One such scale, the Intrinsic Motivation Inventory (IMI) can be used to inform stakeholders of students' subjective motivational and regulatory styles. The IMI is a multidimensional scale developed in support of Self-Determination Theory [1, 2, 3], a strongly validated theory stating that motivation and regulation are moderated by three innate needs: autonomy, belonging, and competence. As applied to education, the theory posits that students who perceive volition in a task, those who report stronger connections with peers and teachers, and those who perceive themselves as competent in a task are more likely to internalize the task and excel. ASSISTments, an online mathematics platform, is hosting a series of randomized controlled trials targeting these needs to promote integrated learning. The present work supports these studies by attempting to validate four subscales of the IMI within ASSISTments. Iterative factor analysis and item reduction techniques are used to optimize the reliability of these subscales and limit the obtrusive nature of future data collection efforts. Such scale validation efforts are valuable because student perceptions can serve as powerful covariates in differentiating effective learning interventions.

Keywords: Intrinsic Motivation Inventory, Self-Determination Theory, ASSISTments, Factor Analysis, Validity, Reliability.

1 Introduction

1.1 Psychometric Research in Online Learning: Value for AIED

Online learning environments allow for the implementation of psychometric scales on diverse samples of students participating in authentic learning tasks. Scales measuring personality traits, values, beliefs, motivation, and other self-reported psychological characteristics, have supported educational research for many years. However, it seems that recent opportunities for data collection at scale, made possible by omnipresent technology, have led many researchers to overlook the procedures necessary to ensure valid measurement.

Although this may not seem like an issue of particular interest to the AIED community, it should be of critical concern. Validating a measure in a learning environment before its formal use strengthens the validity and reliability of resulting claims. AIED researchers commonly focus on advancing models of student learning or affect [4, 5]. Models featuring data collected from a clickstream or sensors can be supplemented by student self-reports from psychometric scales to explain additional variance or reduce error. A recent article in the *Journal of Learning Analytics* highlighted psychometric variables relevant to academic performance including measures of cognitive ability, temperament, personality, motivation, and learning strategies [6]. Although researchers tend to cite published reliability statistics before implementing popular psychometric scales, few employ the exploratory or confirmatory factor analyses (or similar methods) necessary to validate use of the scale in their specific domain, population, and/or learning environment. This is not to say that these techniques are completely foreign to researchers in the field; one positive example observed during a review of related literature established and validated a measure of learners' perceptions of pedagogical agents prior to its use in further research [7]. In contrast, it is common practice in psychology to cite an initial publication as proof of a scale's validation prior to its formal use. As such, the AIED community may benefit from stronger approaches to psychometric application.

A concrete example of the importance of scale validation stems from recent focus in the AIED community toward personalization [8]. Researchers tackle this problem by using learner analytics, data mining techniques, and randomized controlled trials to isolate the most effective learning interventions for each student based on a set of predefined characteristics. In such contexts, self-report measures from psychometric scales can provide an opportunity to explain additional variance between students. Scale scores can be used as dependent measures for the purpose of prediction (i.e., "students with *high prior knowledge* are more likely to report feeling *competent*") or as independent variables for exploring interactions or mechanism (i.e., "students with *high perceptions of autonomy* outperformed those with *low perceptions of autonomy* differentially by treatment condition"). Thus, it is critical to strengthen these metrics by taking steps to validate psychometric scales within specific domains, populations, and/or learning environments.

1.2 Self-Determination Theory and the Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory (IMI) [9] is a multidimensional scale developed in support of Self-Determination Theory (SDT) [1, 2, 3], a strongly validated theory claiming that motivation and regulation are guided by three innate needs: autonomy, belonging, and competence. As applied to education, this theory posits that students who perceive volition in a task, those who report stronger connections with peers and teachers, and those who perceive themselves as competent in the task at hand are more likely to internalize the task and excel. It has been shown that promotion of these needs in educational environments can lead to higher quality learning, as well as greater conceptual understanding, personal growth, and positive adjustment [10]. If validated in an online learning environment, the IMI could potentially be used to inform stakeholders of students' motivational and regulatory styles, alerting them to pertinent implications for learning outcomes and appropriate interventions.

In their landmark review outlining the growth of SDT, Ryan & Deci [2] cite applications of the theory across research domains including education, health care, religion, health and exercise, political activity, environmental activism, and intimate relationships. The IMI has also been applied broadly, with past work validating versions of its subscales in contexts including sports and competition [11, 12], reading [13], mathematics [14], language learning [14], psychiatry [15], medicine [16], puzzle completion [17], computer tasks [18], and teacher training [19]. Past work has also shown IMI subscales to have strong temporal reliability [20]. While examples of IMI application have clearly varied by domain, task, and sample population, it is important to note that they have also varied by scale and item inclusion, scale and item order, and data collection environment. As such, the developers of the IMI encourage researchers to validate the scale within their specific domains, populations, and/or environments of interest [9].

1.3 The Present Work

The present work provides an example of scale validation in an online learning environment using iterative exploratory factor analysis and item reduction techniques. ASSISTments (www.assistments.org), an online learning environment known for its embrace of educational research at scale [21, 22], is currently hosting a series of randomized controlled trials examining learning interventions that target the innate needs defined by SDT with the goal of promoting integrated learning and thereby improving student performance. In support of this research, the present work attempts to validate four subscales of the IMI measuring students' perceptions of autonomy, belonging (or relatedness), competence, and interest/enjoyment within ASSISTments. Validation of IMI subscales within ASSISTments is valuable because students' perceptions can serve as powerful independent or dependent measures when isolating effective learning interventions. Goals of the present work are to achieve convergent, discriminant, and face validity for each subscale, to achieve high reliability for each subscale, and to reduce the number of items within each subscale for future implementation. The latter goal will make future data collection less obtrusive (by requiring fewer items), thereby allowing survey efforts to more easily scale to the broader ASSISTments user population (approximately 50,000 users).

2 Methods

2.1 Sample

Five teachers who regularly work with ASSISTments were contacted with the request that their students participate in a 28-item Likert scale survey. Teachers were notified that the survey would immediately follow a brief assignment (of their choice) used for classwork or homework, and that it would add 5-10 minutes to the assignment based on students' reading levels. Four teachers chose to participate and provided assignments that were modified by the primary author to include two additional items, one introducing the survey as a data collection tool to strengthen students' experiences within ASSISTments and one providing access to the IMI subscales.

Participating teachers and their students were representative of different subpopulations and sampling styles. A total of 226 students participated in at least one of the four subscales. Students of Teacher 1 (n = 73) and Teacher 2 (n = 54) were enrolled in 7^{th} grade math classes at two schools in two different suburban/rural locations in Massachusetts. Teacher 1 chose to embed the IMI subscales after an 8-question homework assignment. Teacher 2 split delivery of the subscales, enrolling her students in a randomized controlled trial including two scales (Interest/Enjoyment and Competence) for homework, and choosing to embed the remaining two scales (Autonomy and Belonging) following an 8-question classwork assignment. These two assignments were strongly conceptually linked and split scale delivery was embraced to examine the potential consequences for reliability and score interpretation within teachers. Students of Teacher 3 (n = 46) were enrolled in high school level math courses in an urban location in Massachusetts and were highly representative of ESL and low SES populations. Teacher 3 chose to embed the subscales after a "class opener" with two multiple choice questions. Students of Teacher 4 (n = 53) were enrolled in high school level engineering courses in an urban location in Massachusetts and represented accelerated learners. Teacher 4 chose to embed the survey following a 20-question assignment on velocity.

All students were familiar with ASSISTments and used the system regularly for classwork and homework in the courses in which they were surveyed. In general, students were not allowed to opt out of survey participation up front but were allowed to skip scale responses and progress to the end of their assignments at any time during their participation. For the RCT-bound scales delivered by Teacher 2, students were prompted to opt-in to survey participation causing unbalanced scale responses between subscales within Teacher 2. This caused average overall missingness (%) to vary by scale across teachers: Interest/Enjoyment (M = 13.60, SD = 0.33), Autonomy (M = 2.16, SD = 0.46), Belonging (M = 3.70, SD = 0.35), and Competence (M = 14.38, SD = 0.37). The analytic sample was reduced based on missing data using listwise deletion. This approach is appropriate for factor analysis because unbalanced items can sway factor loadings [23]. As such, results are based on samples with complete response patterns for modeled scale items.

2.2 Intrinsic Motivation Inventory (IMI)

The IMI is a multidimensional scale intended to measure the subjective experiences of participants following task participation [9, 1]. Various iterations of the IMI have been in use for more than 30 years, with well-established validity and subscale reliability across tasks, conditions, and settings [9]. The scale has six primary subscales that can be mixed and matched to suit research needs: interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice. A seventh subscale intended to measure perceived relatedness or belonging was added in recent years and has not yet been established as valid or reliable.

Subscales. All scale items were modified slightly to reflect an academic task or setting; such modifications are thought to be inconsequential to outcomes [9]. Students were asked to indicate how true each statement was for them using a Likert scale (1 = *Not at All True*, 7 = *Very True*). Past work has suggested that order effects of scale and item delivery are negligible and that subscales can be included or excluded as necessary [24; 9]. The four subscales considered in the present work align with the basic psychological needs defined by Self-Determination Theory, as detailed in the subsections below.

Interest/Enjoyment. This subscale is the primary measure of intrinsic motivation. It includes seven items regarding intrinsic motivation (i.e., "I enjoyed doing this assignment very much"), with two items reverse scored (i.e., "This assignment did not hold my attention at all").

Autonomy. This subscale is the primary measure of perceived autonomy, also known as choice, volition, or task-based locus of control. Scores on this scale have previously been shown to predict Interest/Enjoyment scores [9]. This subscale includes seven items regarding perceived autonomy (i.e., "I did this assignment because I wanted to"), with five items reverse scored (i.e., "I did this assignment because I had to").

Belonging. This subscale is the primary measure of perceived relatedness or belonging. This scale was added to the IMI in recent years and does not have well established validity or reliability. In addition, modifications to items in this subscale to capture how well students felt they related to their classmates may have been more significant than modifications to other scales because the effect was extrapolated to a collective group (i.e., changing "task" or "activity" to "assignment" does not extrapolate to *many tasks*). This subscale includes eight items (i.e., "I'd like a chance to interact with my classmates more often") with four items reverse scored (i.e., "I don't feel like I could really trust my classmates").

Competence. This subscale is the primary measure of perceived competence or feeling capable and confident. Scores on this scale have previously been shown to predict Interest/Enjoyment scores [9]. This subscale includes six items (i.e., "I am satisfied with my performance on this assignment") with one item reverse scored (i.e., "This was an assignment that I couldn't do very well").

2.3 Procedure

Data was retrieved by integrating Qualtrics, a readily accessible survey infrastructure, with ASSISTments using the ASSISTments Survey System available through the ASSISTments TestBed [25]. This system uses an iframe to establish a connection between the two platforms, resulting in the ability to link survey data to ASSISTments performance through anonymized student and assignment identification numbers. Two items were added to the end of each participating teacher's assignment: a verification item introducing the survey as a data collection tool to strengthen students' experiences within ASSISTments, and an item with an embedded iframe that connected students to the survey content in Qualtrics while they worked in ASSISTments.

IMI items were delivered through Qualtrics using subscale alignment. Except for those of Teacher 2, all students were asked to respond to all items pertaining to Interest/Enjoyment in a single page view. When finished, or if opting not to answer, they could select "Next" to move on to the next page and subscale. Students cycled through the Autonomy, Belonging, and Competence subscales in this fashion until ultimately completing the survey. Due to Teacher 2's split delivery, her students opted-in to the Interest/Enjoyment and Competence subscales after a homework assignment and received the Autonomy and Belonging subscales after a subsequent class assignment using the same protocol and infrastructure noted above.

The data collection period lasted one week. Data was retrieved from Qualtrics and ASSISTments, compiled, and preprocessed for IBM SPSS Statistics. Variables were cleaned, and missing data was labelled for proper exclusion from analysis. Redundancies were removed while merging data (24 students accessed the survey multiple times; in these cases, only first responses were retained). The resulting data file contained responses from 226 students. Items were reverse scored as necessary - a step not required for factor analysis, but critical for calculating reliability using Cronbach's α [23]. Higher score values indicated higher levels of scale sentiment across all scales (i.e., greater enjoyment). De-identified survey data is available at [26] for additional reference.

Following the guidelines set forth by Field [23] iterative scale reduction was conducted using principal axis factor analyses. Given the likelihood of correlations between subscales, oblique rotation (i.e., direct oblimin) offered a more appropriate approach than orthogonal rotation (i.e., varimax). Factors were established using traditional methods: factors with eigenvalues greater than 1.0 were considered valid for inclusion using Kaiser's criterion, and scree plots were developed to confirm factor count by estimating the point of inflexion. Items were removed as part of the iterative process to establish stronger validity and reliability. Inter-item correlations and subscale reliability measures were consulted for scale reduction. Where items were removed, factor analysis was repeated to assess potential changes to factors and loadings and to optimize the model.

3 Results

3.1 Iterative Scale Reduction

28-Item Factor Analysis. A principal axis factor analysis was conducted on all 28 items from the IMI subscales using direct oblimin oblique rotation. After listwise deletion of missing responses the analytic sample consisted of 180 students. The Kaiser-Meyer-Olkin measure verified the sample was large enough for analysis, KMO = .81, and Bartlett's Test of Sphericity was significant, χ^2 (378) = 2,818.75, $p < .001$ (as desired). In addition, the diagonals of the anti-image correlation matrix were above .5 (as desired). Thus, 180 students provided an adequate sample size for analysis.

This analysis was conducted to assess initial model structure and examine the potential for item reduction. The model had poor structure, as suggested by a determinant of 5.61E-008 (denoting issues of multicollinearity), numerous after extraction communalities below 0.70, and an average communality of 0.58 denoting that Kaiser's criterion was not necessarily an appropriate threshold for factor inclusion. The model resolved to six factors using Kaiser's criterion and four or six factors based on interpretation of the scree plot's point of inflexion. Four factors, as desired based on the initial subscales, accounted for 56% of the variance in the model, with the remaining two factors accounting for an additional 11%.

The correlation matrix attained from this analysis (Table 1) was helpful in reducing scale items to establish a stronger model. Field [23] suggests beginning the reduction process by assessing the correlation matrix for multiple inter-item correlations over .90 or under .30. As none of the correlations exceeded .90, items were slated for removal from the model if 50% or more of the within-scale inter-item correlations were .30 or less (i.e., item E4 had 3/6 correlations of .30 or smaller; suppressed correlations are all less than .30). Using this approach, eight items were removed from the model: Interest/Enjoyment Item 4 (50%), Autonomy Item 6 (50%), Belonging Items 2 (57%), 3 (57%), 4 (86%), 5 (57%), and 7 (57%), and Competence Item 6 (100%).

20-Item Factor Analysis. Following removal of these eight items, a second principal axis factor analysis was conducted on the remaining 20 items, again using direct oblimin oblique rotation. The analytic sample again consisted of an adequate sample size of 180 students: the Kaiser-Meyer-Olkin measure verified the sample was large enough for the analysis, KMO = .84, and Bartlett's Test of Sphericity was significant, χ^2 (190) = 2,066.79, $p < .001$.

Model structure was still not ideal. The expected structure was resolved, with four factors retained (using Kaiser's criterion), explaining 66% of the variance and with all items loading on their expected subscales. However, although the determinant increased to 5.84E-006, it remained lower than the desired minimum of 1.00E-005, suggesting that a multicollinearity issue remained and that additional items should be considered for removal.

Having addressed the issue through the correlation matrix, no additional reductions were suggested using this approach. Thus, reliability analyses were conducted on each subscale to determine candidates for removal. Cronbach's α was used with listwise deletion of missing values by scale. Analysis suggested that reliability of the six remaining items in the Interest/Enjoyment subscale was high, $\alpha = .91$ ($n = 191$), but could be increased by removing Item 3. Reliability for other subscales was mixed, as shown in Table 2, but no other items met qualifications for removal using this approach. Therefore, Interest/Enjoyment Item 3 was removed from the model, leaving 19 items.

19-Item Factor Analysis. Following item removal, a final principal axis factor analysis was conducted on the remaining 19 items, again using direct oblimin oblique rotation. Listwise deletion of missing data left an analytic sample of 181 students. The Kaiser-Meyer-Olkin measure verified the sample was large enough for the analysis, KMO = .83, and Bartlett's Test of Sphericity was significant, χ^2 (171) = 1,988.72, $p < .001$.

Table 1. Inter-item correlations in the 28-item model (n = 180).

	Interest/Enjoyment							*Autonomy*							*Belonging*								*Competence*					
Scale/Item	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	1	2	3	4	5	6
Enjoyment																												
E1	--																											
E2	**.83**	--																										
E3	**.52**	**.53**	--																									
E4	**.32**	**.33**	**.62**	--																								
E5	**.69**	**.67**	**.43**	**.19**	--																							
E6	**.75**	**.77**	**.47**	**.30**	**.74**	--																						
E7	**.62**	**.64**	**.34**		**.70**	**.65**	--																					
Autonomy																												
A1	**.42**	**.37**	**.36**	**.21**	**.41**	**.35**	**.36**	--																				
A2	.15		**.24**	**.25**	.18		.14	**.39**	--																			
A3	.17	.17	**.32**	**.24**	**.21**	.17		**.43**	**.53**	--																		
A4	.13	.15	**.20**	.17	**.24**	.15	**.18**	**.31**	**.39**	**.53**	--																	
A5	**.24**	**.20**	**.30**	**.25**	**.30**	**.31**	**.23**	**.51**	**.42**	**.57**	**.48**	--																
A6	**.47**	**.52**	**.37**	**.21**	**.56**	**.43**	**.44**	**.45**	**.24**	**.31**	**.21**	**.45**	--															
A7	**.18**		**.23**	.17	**.26**	**.22**	**.22**	**.39**	**.36**	**.40**	**.58**	**.59**	**.24**	--														
Belonging																												
B1	-.13				-.14	-.14								.15	--													
B2													-.13	.14	**.67**	--												
B3							.15								**.23**	**.18**	--											
B4																	**.25**	--										
B5					**-.23**	-.14	-.16	**-.23**					**-.23**		**.45**	**.39**	**.26**		--									
B6										.15					**.43**	**.35**	**.38**		**.47**	--								
B7														**-.18**	.16		**.40**	**.44**	**.26**		--							
B8					.13		**.19**								**.35**	**.24**	**.64**	**.18**	**.27**	**.31**	**.47**	--						
Competence																												
C1									**-.24**	**-.18**	**-.19**	-.14			**.24**	.13	**.19**		**.18**			.16	--					
C2									**-.21**	-.16	**-.23**	**-.25**		**-.21**	**.18**	.13	**.27**				.14	**.19**	**.61**	--				
C3	.13	.17			.13		.15		**-.18**	-.13	-.15		.16	-.15	.17		**.33**				**.21**	**.26**	**.62**	**.56**	--			
C4									-.16	-.13	-.15				**.34**	**.22**	**.21**		.17		**.20**	.17	**.77**	**.61**	**.66**	--		
C5			-.14					-.12	**-.29**	**-.25**	**-.18**	**-.22**			**.28**	.15	**.29**		**.18**		**.23**	**.25**	**.78**	**.70**	**.68**	**.83**	--	
C6	**-.25**	-.15	**-.21**		**-.26**	-.15	**-.18**	**-.33**	**-.27**	-.16		**-.23**	**-.27**		**.22**	**.31**	.15		**.29**	**.24**			.15	.14	.17	**.19**	**.29**	--

*Note***.** Bold correlations, p < .01; all others, p < .05. Suppressed correlations were not significant at p < .05.

Table 2. Reliability of subscales in the 20-item model and candidate items for removal.

Scale	*n*	*Scale Items*	α	α *if item removed*
Interest/Enjoyment	191	6	91	.92 (Item 3)
Autonomy	220	6	.83	--
Belonging	217	3	.56*	--
Competence	192	5	.92	--

*The Belonging subscale does not have well-established validity or reliability; low reliability of this scale was not of immediate concern.

This final iteration established adequate model structure. Four factors were retained using Kaiser's criterion, explaining 68% of the variance, with all items loading on their expected subscales. Four factors were also suggested via scree plot interpretation, as shown in Figure 1. The determinant increased to 1.01E-005, just surpassing the threshold of 1.00E-005 and suggesting that multicollinearity had been sufficiently resolved. Average communality increased to 0.595, bordering the 0.60 threshold for adequacy of the Kaiser criterion. Additionally, only 12% of the residuals in the reproduced matrix had absolute values greater than 0.05. Reliability results were unchanged from those presented in Table 2, with the noted increase in reliability for the Interest/Enjoyment subscale following item removal. Further, all corrected item-total correlations were above 0.30, suggesting that each item correlated well with its respective overall scale score. Values were lowest for items in the Belonging subscale, suggesting issues with the validity of using this subscale alongside others in the IMI. Overall reliability of all 19 items in the model was moderate, α = .78.

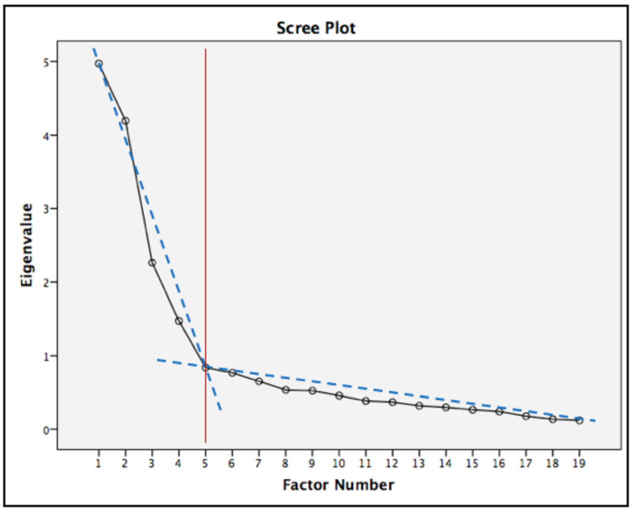

Fig. 1. Scree plot with point of inflexion mapped. The red vertical line at factor 5 designates the non-inclusive cutoff factor, or the 'elbow' of the graph.

Table 3. Summary of exploratory factor analysis results in the 19-item model (n = 181).

Scale/Item	*Interest/Enjoyment*	*Competence*	*Autonomy*	*Belonging*
Interest/Enjoyment				
IE2 – This assignment was fun to do.	.879			
IE1 – I enjoyed doing this assignment very much.	.858			
IE6 – I thought this assignment was quite enjoyable.	.857			
IE5 – I would describe this assignment as very interesting.	.810			
IE7 – While I was doing this assignment, I was thinking about how much I enjoyed it.	.767			
Competence				
C4 – I am satisfied with my performance on this assignment.		.935		
C5 – I was pretty skilled at this assignment.		.915		
C1 – I think I am pretty good at this assignment.		.868		
C3 – After working at this assignment for a while, I felt pretty competent.		.709		
C2 – I think I did pretty well at this assignment, compared to other students.		.664		
Autonomy				
A5 – I did this assignment because I had no choice. (R)			.780	
A3 – I didn't really have a choice about doing this assignment. (R)			.720	
A7 – I did this assignment because I had to. (R)			.704	
A4 – I felt like I had to do this assignment. (R)			.663	
A2 – I felt like it was not my own choice to do this assignment. (R)			.581	
A1 – I believe I had some choice about doing this assignment.			.518	
Belonging				
B6 – I'd really prefer not to interact with my classmates in the future. (R)				.665
B1 – I feel really distant to my classmates. (R)				.643
B8 – It is likely that my classmates and I could become friends if we interacted a lot.				.492
Eigenvalues	4.97	4.19	2.27	1.47
% of variance	26.16	22.07	11.92	7.74
α	.92	.92	.83	.56

3.2 Resulting Subscales in the 19-Item Model

The four resulting factors aligned with expected subscales. Table 3 provides each scale item with factor loadings from the pattern matrix after rotation. The pattern matrix was used for interpretation because it ignores shared variance and shows the unique contribution of items to factors [23]. Factor 1 aligned with the Interest/Enjoyment subscale, explaining 26.14% of model variance. Similarly, Factor 2 aligned with the Competence subscale (22.07%), Factor 3 aligned with the Autonomy subscale (11.92%), and Factor 4 aligned with the Belonging subscale (7.74%). The Interest/Enjoyment, Autonomy, and Competence subscales showed substantial convergent validity and high reliability, and all scales displayed high discriminant and face validity.

3.3 Class Variations

A brief investigation was conducted to examine how scale scores using the 19-item model varied across participating teachers. Split-file analysis was used to assess the reliability of each subscale using Teacher as the grouping variable. Results are shown in Table 4. Subscales showed similar patterns of reliability regardless of teacher, with some variation in magnitude across teachers. Of note, Teacher 2 and Teacher 3 exhibited the lowest reliability on the Belonging subscale, while Teacher 2 and Teacher 4 exhibited lower than anticipated reliability on the Competence subscale.

Table 4. Subscale reliability by teacher.

	Teacher 1		Teacher 2		Teacher 3		Teacher 4	
Scale	*n*	α	*n*	α	*n*	α	*n*	α
Interest/Enjoyment	73	.93	27	.88	38	.92	53	.93
Autonomy	71	.81	54	.77	41	.78	53	.88
Belonging	71	.71	54	.29	39	.23	53	.75
Competence	70	.93	29	.82	40	.90	53	.82

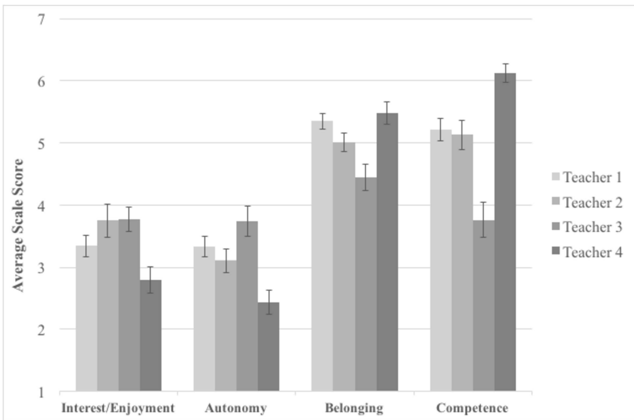

Fig. 2. Resulting scale scores by teacher.

Scale scores were defined for each student by averaging factor items and then aggregated by teacher for final comparison (see Figure 2). Such aggregates offer an example of how this psychometric scale could be used to establish useful variables or covariates for future research. ANOVAs revealed significant differences between teachers on all subscales: Interest/Enjoyment, $F(3, 190) = 3.71$, $p < .05$; Autonomy, $F(3, 218) = 6.98$, $p < .001$; Belonging, $F(3, 216) = 6.76$, $p < .001$; and Competence, $F(3, 191) = 21.08$, $p < .001$. Given numerous confounds in present survey collection (e.g., teacher, assignment, skill level, age range) further assessment was not considered. However, future work could control for potential sources of variance to better define the mechanisms underlying these significant differences.

4 Contributions & Limitations

The goal of the present work was to validate subscales of the IMI within ASSISTments to support a series of randomized controlled trials assessing the efficacy of learning interventions that target students' perceptions of autonomy, belonging, and competence. Employing an iterative factor analysis and item reduction approach with an analytic sample of 181 students established substantial convergent validity and high reliability for three reduced IMI subscales (Interest/Enjoyment, Autonomy, and Competence). Issues were observed with the reliability of the Belonging subscale, due in part to the high proportion of items removed to optimize the model. This subscale is not well-established as valid or reliable [9] and it did not perform well within ASSISTments. As such, the Belonging subscale will not be used in future data collection efforts and students' aggregate scores on this subscale will not be used in future analyses. Results also suggested that all four subscales exhibited high discriminant and face validity.

Limitations of this work include potential bias introduced by varied delivery protocol of subscales (i.e., Teacher 2's split delivery), delivery of items aligned within subscales (although previously addressed as inconsequential for psychometric scales [9, 24]), and the potential for reduced external validity due to item reduction. The randomized controlled trials supported by this work will use the subscales established by the 19-item model when compiling aggregate scores for use in future analyses. The reduction of 9 items from these subscales will also make future data collection using the IMI within ASSISTments less obtrusive, thereby allowing survey efforts to scale. With hope, these results also serve as a valuable reminder for the AIED community that contextually validating a psychometric scale prior to its formal use strengthens the validity and reliability of resulting claims.

Acknowledgments

We acknowledge funding from NSF grants (ACI-1440753, DRL-1252297, DRL-1109483, DRL-1316736, DGE-1535428, OAC-1724889, OAC-1636782 & DRL-1031398), the U.S. Department of Education (IES R305A120125 & R305C100024 and GAANN), the ONR, and the Gates Foundation. Thanks to S.O. & L.P.B.O.

References

1. Ryan, R.: Control and information in the intrapersonal sphere an extension of cognitive evaluation theory. *Journal of Personality and Social Psychology,* **43**, 450-461 (1982).
2. Ryan, R. M., Deci, E. L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist,* **55**, 68-78 (2000).
3. Ryan, R. M., Deci, E. L.: An overview of self-determination theory. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of Self-Determination Research.* Rochester, NY: University of Rochester Press (2002).
4. Brusilovsky, P., Millan, E.: User models for adaptive hypermedia and adaptive educational systems. In Brusilovsky, Kobsa, & Nejdl (Eds.) *The Adaptive Web: Methods and Strategies of Web Personalization.* LNCS 4321, Springer-Verlag Berlin Heidelberg (2014).
5. Calvo, R. A., D'Mello, S.: Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transaction on Affective Computing,* **1** (1), 18-37 (2010).
6. Gray, G., McGuinness, C., Owende, P., Carthy, A.: A review of psychometric data analysis and application in modelling of academic achievement in tertiary education. *Journal of Learning Analytics,* **1** (1), 75-106 (2014).
7. Ryu, J., Baylor, A. L.: The psychometric structure of pedagogical agent persona. *Tech. Inst. Cognition and Learning,* **22**, 291-314 (2005).
8. Santos, O. C., Kravcik, M., Boticario, J. G.: Preface to special issue on user modelling to support personalization in enhanced educational settings. *International Journal of Artificial Intelligence in Education,* **26** (3), 809-820 (2016).
9. Intrinsic Motivation Inventory, http://selfdeterminationtheory.org/questionnaires/, last accessed 2018/02/07.
10. Deci, E., Vallerand, R., Pelletier, L., Ryan, R.: Motivation and education: The self-determination perspective. *Educational Psychologist*, **26**, 325-346 (1991).
11. McAuley, E., Duncan, T., Tammen, V. V.: Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport* 60, 48-58 (1987).
12. McAuley, E., Tammen, V. V.: The effects of subjective and objective competitive outcomes on intrinsic motivation. *Journal of Sport and Exercise Psychology,* **11**, 84-93 (1989).
13. Grolnick, W., Ryan, R.: Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, **52**, 890-898 (1987).
14. Monteiro, V., Mata, L., Peixoto, F.: Intrinsic Motivation Inventory: psychometric properties in the context of first language and mathematics learning. *Psicologia: Reflexao e Critica,* **28** (3), 434-443 (2015).
15. Choi, J., Mogami, T., Medalia, A.: Intrinsic Motivation Inventory: An adapted measure for schizophrenia research. *Schizophrenia Bulletin,* **36** (5), 966-976 (2010).
16. Williams, G. C., Deci, E. L.: Supporting autonomy to motivate glucose control in patients with diabetes. *Diabetes Care,* **21**, 1644-1651 (1998).
17. Ryan, R., Mims, V., Koestner, R.: Relation of reward and interpersonal context to intrinsic motivation: A review and test using Cognitive Evaluation Theory. *Journal of Personality and Social Psychology*, **45**, 736-750 (1983).
18. Deci, E., Eghrari, H., Patrick, B., Leone, D.: Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, **62**, 119-142 (1994).
19. Filak, V., Sheldon, K.: Student psychological need satisfaction and college teacher-course evaluation. *Educational Psychology*, **23**, 235-247 (2003).
20. Tsigilis, N., Theodosiou, A.: Temporal stability of the Intrinsic Motivation Inventory. *Perceptual and Motor Skills,* **97** (1), 271-280 (2003).

21. Heffernan, N., Heffernan, C.: The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education,* **24** (4), 470-497 (2014).
22. Ostrow, K. S., Heffernan, N. T., Williams, J. J.: Tomorrow's edtech today: establishing a learning platform as a collaborative research tool for sound science. *Teachers College Record,* **119** (3), 1-36 (2017).
23. Field, A.: *Discovering statistics using IBM SPSS Statistics*. 4th edn. SAGE Publications, London (2013).
24. Schell, K. L., Oswald, F. L.: Item grouping and item randomization in personality measurement. *Personality and Individual Differences,* **55**, 317-321 (2013).
25. ASSISTments TestBed Resource Guide, https://www.assistmentstestbed.org, last accessed 2018/02/06.
26. Submission Data, https://osf.io/4gwuc/, last accessed 2018/02/07.