International Journal of Education and Human Developments, Vol. 10 No. 1; November 2024
ISSN 2415-1270 (Online), ISSN 2415-1424 (Print)
Published by Center for Global Research Development

Development and Validation of Computer Game Engagement Instrument

Sunha Kim, PhD
Associate Professor
Department of Counseling, School, and Educational Psychology
Department of Learning and Instruction
Graduate School of Education
SUNY at Buffalo
423 Baldy Hall, Buffalo, NY 14260, USA.

Mido Chang, PhD
Professor
Department of Counseling, Recreation, & School Psychology
College of Arts, Sciences & Education
Florida International University
11200 SW 8th St. ZEB 250B, Miami, FL 33199, USA.

Hyejin Bang, PhD
Professor
Department of Counseling, Recreation, & School Psychology
College of Arts, Sciences & Education
Florida International University, USA.

Abstract

Despite the growing interest in computer games as instructional media to enhance students' learning in school, there remains a lack of reliable and valid measurement tools to evaluate their effectiveness. In response to this gap, we have developed the Computer Game Engagement Instrument (CGEI). This instrument aims to assess students' levels of engagement in instructional games, which research has identified as a crucial factor influencing academic engagement. This study follows Messick's framework for the development and validation of the CGEI and utilizes the Rasch measurement model. We administered the CGEI to 250 postsecondary school students to analyze items and test characteristics, as well as the test-takers' ability levels. The study explores the psychometric properties of the CGEI and discusses its implications for educational practice. Our findings provide empirical evidence supporting the validity of the CGEI, offering a valuable tool for evaluating the engagement levels of students in instructional computer games.

Keywords: Computer games, computer engagement. Messick's validity, Rasch measurement theory, instrument development

1. INTRODUCTION

As influential educational tools, computer games have garnered the attention of numerous educational researchers and practitioners seeking to assess their efficacy in promoting student learning (Annetta et al., 2008; Ke, 2008; Papastergiou, 2009). However, there is a scarcity of instructional computer games proven to be effective in enhancing student learning outcomes.

This dearth of successful outcomes may stem from the absence of valid and reliable measures to evaluate the effectiveness of instructional computer games, as highlighted by Betts (2012), who underscores the importance of measurement as the foundation for research.

In this study, we introduce a newly developed computer game instrument that incorporates robust psychometric properties. Specifically, our focus lies on students' engagement in computer games as a central construct of the measure. The motivation behind studying student engagement stems from its recognized significance as a precursor to performance (Lau & Roeser, 2002). Moreover, engagement is a pivotal individual factor influencing students' propensity to engage with computer games (Funk, 2002). Thus, the objective of this study is to develop and validate an instrument, the Computer Game Engagement Instrument (CGEI), designed to measure students' engagement in game-mediated learning. For the development and validation of the CGEI, we adopted Messick's validity framework (1995). Additionally, this study draws upon a comprehensive literature review to establish operational definitions of game engagement and its three subcomponents.

In an effort to provide empirical evidence for the validity of the CGEI using Messick's framework, this study underwent two stages of the process. Initially, we conducted a pilot test involving 50 postsecondary school students. The data was analyzed for internal consistency of items using jMetrik 1.0.3, following the methodology employed by Meyer (2009).

Subsequently, a field test was conducted, collecting data from 250 postsecondary school students. We employed the Rasch measurement model (Rasch, 1961) to assess the validity of the CGEI.

2. BACKGROUNDS & METHODS

2.1. Messick's Framework

Messick's framework (1995) encompasses six components: Content, Substantive, Structural, Generalizability, External, and Consequential. The content component focuses on the relevance and representativeness of item content to the target construct, alongside its technical (statistical) quality. The substantive component examines whether respondents engage with the assessment process as intended by the instrument developer. The structural component evaluates the extent to which item interrelationships align with the theoretical underpinnings of the construct. Generalizability pertains to the extent to which instrument features and interpretations are applicable across diverse populations, tasks, and contexts. The external component assesses the degree to which the study construct correlates with other external variables. Notably, this study did not incorporate the consequential component, which evaluates the long-term effects of the instrument's administration on individuals or systems, as the research did not aim to assess the value or impact of the instrument.

Throughout the development and validation of the CGEI, Messick's framework was also extensively applied. During the instrument development stage, we delineated the purpose of the CGEI, provided a theoretical framework, specified test structure, and developed specific items. Clearly defining the instrument's purpose and rationale was crucial for content validation. Presenting theoretical support with precise operational definitions of the construct served as evidence for substantive validity. Additionally, grounding the internal model on reliable theories provided evidence for both substantive and structural validity.

During the instrument validation stage, we provided and evaluated empirical-data-based evidence for validity by fitting the Partial Credit Model of Rasch Measurement Theory using WinSteps 3.58 (Linacre, 2005). This involved reviewing technical qualities of items such as item difficulty and discrimination to establish content validity. For the substantive aspect, we analyzed person fit and item hierarchy while constructing a rating scale model. Overall, this study presented dimensionality for structural validity, reliability for generalizability, and item-person maps for external validity.

2.2. Stages for Instrument Development

During the development stage, our primary focus was on defining game engagement as the central construct of the instrument and constructing its internal model. This process was meticulously designed to underpin the content, substantive, and structural aspects of validity. For our internal model, we adopted the engagement construct by Fredricks et al. (2004).

Engagement is a multifaceted construct that can be subdivided into three components: behavioral, cognitive, and emotional engagement. Although there is no universally agreed-upon definition of engagement, Fredricks et al. (2004) underscore commitment or investment as key aspects of engagement. Behavioral engagement encompasses participation, involving involvement in tasks, completing required work, and adhering to rules. Cognitive engagement involves mindfulness and a willingness to exert effort to comprehend complex ideas and master advanced skills. Emotional engagement pertains to affective reactions, including interest, happiness, and a sense of belonging. To ensure content validity, we included items that represent the three components of engagement in the internal model. We expected that measures of behavioral, emotional, and cognitive engagement would be correlated with each other, anticipating strong internal consistency and a high reliability coefficient during item analysis of the instrument.

For test specifications, we drew from existing literature on the definitions of the three components of engagement (Appleton et al., 2006; Fredricks et al., 2004; Skinner, Furrer et al., 2008; Patrick et al., 2007), and formulated operational definitions for each component. Subsequently, items for each component were created based on these operational definitions, ensuring content validity by representing all sub-domains of engagement. To establish construct validity in the internal structure, this study conducted a principal component analysis (PCA) during the validation stage. It was anticipated that these items would measure each of the three subcomponents, converging into a major overarching construct of game engagement. This approach aligns with the perspectives of other researchers (Perry et al., 2010; Simons-Morton & Chen, 2009) and corresponds to the assumption of Rasch's unidimensional measurement model (Bond & Fox, 2001).

The instrument comprised 34 polytomous items of engagement, which were recoded on a six-point Likert scale: 1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=somewhat agree, 5=agree, 6=strongly agree. Additionally, the items featured a "Not Applicable (N/A)" option. The ordinal nature of the items facilitated the assignment of numeric values, contributing to substantive validity. By averaging the scores of individuals, with numeric values ranging from 1 to 6, it became possible to ascertain whether an individual demonstrated a high or low level of engagement in learning via computer games.

After conceptualizing the instrument, we developed a test blueprint to provide further guidance for instrument development. This blueprint comprised two dimensions: content classifications and process levels. In the content classification, each item was categorized into three subdomains of engagement: behavioral, emotional, and cognitive engagement. Concerning process levels, three levels were utilized based on Bloom's affective taxonomy (Krathwohl et al., 1984): receiving (being aware of one's engagement status in playing computer games, e.g., "I felt happy when playing the computer game"), responding (being willing to engage in issues, e.g., "I was willing to tackle challenging tasks in the computer game"), and valuing (forming a judgment about playing computer games, e.g., "The computer game was beneficial for improving thinking skills").

This study focused on the three low levels of Bloom's taxonomy, omitting the two high levels of organization and internalizing value, as these require the establishment of a value system regarding educational games, which lies beyond the scope of this study. Table 1 provides an overview of the test specifications, indicating the number of items in the CGEI belonging to each of the two dimensions: content and process level.

Table 1. **Test Specification for CGEI**

	Content		
Level	Behavior Engagement	Emotional Engagement	Cognitive Engagement
1. Receiving	1	7	5
2. Responding	10	1	6
3. Valuing		4	

Following the test specifications and blueprints, an instrument measuring game engagement was developed. The instrument comprised two parts: Part 1 collected background information of participants, while Part II consisted of 34 items measuring participants' experience of playing computer games, specifically evaluating their engagement levels. This study conducted content classification and process level analysis for each item.

2.3. Stage for Instrument Validation

A total of 250 graduate students participated in the main data collection and field test. Prior to the main analysis, a pilot test was conducted with 50 graduate students. The pilot test data were analyzed using jMetrik to explore the internal consistency of items and identify any issues with the instrument.

For the main analysis of data from the 250 college students, this study adopted the Rasch model. The Rasch measurement theory identifies an ideal probability model and selects items that fit the identified model. The probability of answering test items correctly depends on both respondents' ability levels and item difficulty. Specifically, the Partial Credit Model (PCM) was adopted for the analysis of polytomous items. The PCM is suitable for instruments with items whose response choices range from low to high, indicating a higher score as higher ability (Embreston & Reise, 2000).

As part of the validation analysis plan, this study included item/person fit, item quality, item hierarchy, rating scale analysis, dimensionality, reliability, and person-item maps using WinSteps (Linacre, 2005). Item fit analysis was conducted to determine the fit of items, with further examination of empirical Item Characteristics Curves (ICC) for misfit items. Item analysis was then conducted to explore item difficulty and discrimination. Person fit analysis was conducted to identify misfitting individuals, with a check on Z-residual status for systematic errors. Item hierarchy was reviewed based on internal structure, and Rating Scale Analyses were conducted comparing Rating Scale Model (RSM) with Partial Credit Model (PCM). Dimensionality was investigated through Principal Components Analysis of residuals (PCA), and reliability was assessed using Cronbach's Alpha, Rasch reliability, and Person Separation. Test Information Function was examined to identify where the instrument was most accurate, and Item Person map was explored to detect the target of the CGEI instrument.

3. RESULTS

3.1. Validation Results from Piloting, Item Fit, and Person Fit

The pilot test results using jMetrik indicated a relatively high reliability ($\alpha=0.89$). Out of 51 responses, 40 were deemed valid for the pilot testing analysis. The item analysis showed a relatively high coefficient ($\alpha=0.89$). Item means ranged from 2.48 to 5.05, with no extreme locations observed across the 1 to 6 scales. Most items demonstrated a positive correlation with the total score.

Using data from 250 participants, item fit analyses were conducted to assess how well the data fitted into the Rasch model and to identify any unexpected response patterns. Item fit analysis supports content validity by examining the statistical quality of items and their fit with the measurement model. Outfit standardized Z value (ZSTD) was utilized to determine item fit (Smith, 2000), with values greater than positive 1.96 indicating misfits. Nine items were flagged as misfits.

To investigate the flagged items, empirical Item Characteristics Curves (ICC) were examined for each of the nine flagged items (Linacre, 2005). The ICCs indicated misfits occurring at both low and high ends, suggesting low discrimination among individuals at low and high ability levels. Some individuals with high ability levels ($1.96<\theta<5$) tended to score lower than their abilities, while others with low ability levels ($-1.96<\theta\leq-1$) tended to score higher than their abilities. Further examination of these flagged items revealed no extreme difficulty levels and positive correlations with the total score.

Person fit analysis supports substantive validity by assessing whether individuals' responses to items align with expectations. Using Winsteps, person fit was examined, with values greater than 1.96 positive Outfit ZSTD indicating misfits. Out of 250 respondents, 45 individuals were identified as misfits, accounting for 18% of the sample.

To investigate potential systematic measurement errors, the Z-residual status of misfit individuals was examined. Results indicated that misfits occurred randomly across the 34 items, rather than systematically. Many misfit results might be attributed to the respondents' diverse experiences with various types of games.

3.2. Item Hierarchy and Rating Scale Analysis

Item hierarchy was reviewed by analyzing item difficulty levels across the three subcomponents of engagement. WinSteps results revealed minimal differences in difficulty levels across subcomponents. As this study did not assume any distinction in item hierarchy among the three subcomponents, the observed lack of difference in item difficulties aligned with the expected outcomes during instrument development, thereby establishing substantive validity.

Rating Scale Analysis Further analysis was conducted to determine whether the CGEI could utilize the Rating Scale Model (RSM), assuming equivalence across four scales. Using Winsteps, both RSM and Partial Credit Model (PCM) were run, and the relative fits of the models were compared (see Table 2). According to the Chi-Square test result, a significant difference between RSM and PCM fit statistics was observed: the difference (X2 Difference = 326.34, DF= 93) exceeded the critical value of 128 with 93 degrees of freedom at a 0.01 confidence level. PCM fitted significantly better than RSM.

Table 2. Model Comparison: RSM vs. PCM

Model	**Chi-square**	**df**
RSM	22,231.57	8,016
PCM	21,905.23	7,923
X^2 Difference	326.34	93

3.3. Dimensionality

Dimensionality analysis supports the structural aspect of validity. It examines whether the unexplained variance is significant after utilizing the first dimension, or if the dimensional structure suggested by the data aligns with the internal model. To investigate dimensionality, this study conducted Principal Components Analysis (PCA), assuming that the instrument measures only one underlying construct as per the Rasch measurement model. Similarly, this study presumed that the CGEI measures only one construct of game engagement, consistent with other researchers' approaches of unidimensional combinations of subcomponents (Perry, Liu, & Pabian, 2010; Simons-Morton, & Chen, 2009).

Table 3 presents the results of PCA, and Table 4 outlines the six rules used to determine unidimensionality (Linacre, 2006). The results indicated one dominant Rasch dimension, explaining 36.5% of the total variance with an Eigenvalue of 19.6. While the first factor (Rasch dimension) did not account for more than 50% of the total variance, it explained a substantial proportion compared to other factors, which explained less than approximately 6%. The Eigenvalues of these other factors were around or below 3. Additionally, the potential second dimension was significantly lower than the major Rasch dimension with an Eigenvalue of 19.6.

Table 3. **PCA of Residuals**

Scale	# of items	Component	Eigenvalue	Total % Var. Explained	Number of loadings > |.30|
CGEI	34	Rasch	19.6	36.5%	
		1st residual	3.3	6.1%	14
		2nd residual	3.0	5.6%	12
		3rd residual	2.5	4.7%	11
		4th residual	2.1	3.9%	11

***Table 4.* Unidimentionality Rules**

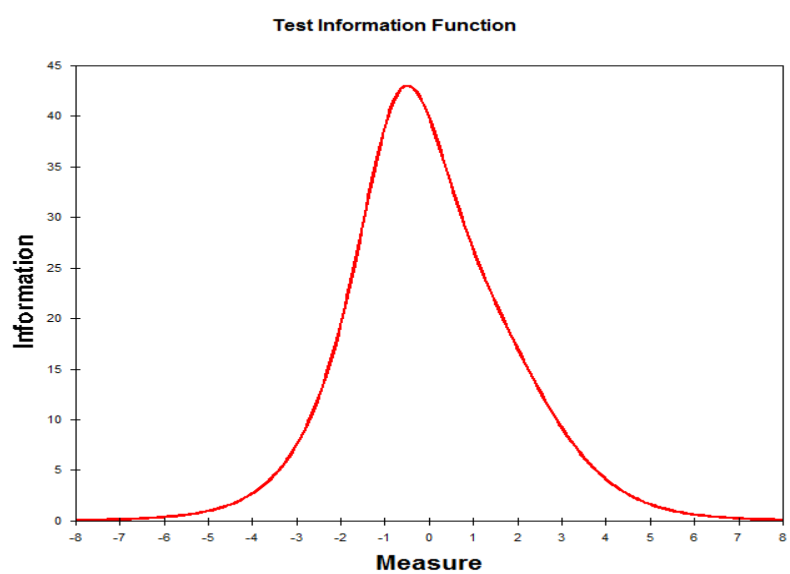
Unidimentionaity Rule	Rule Abidance
Variance explained by measures > 4 x 1st Contrast is good.	Yes
Variance explained by measures > 10 x 1st Contrast is excellent.	No
Variance explained by measures > 50% is good.	No
Unexplained variance explained by 1st contrast (eigenvalue) < 3.0 is good.	Slightly Yes
Unexplained variance explained by 1st contrast (eigenvalue) < 1.5 is excellent.	No
Unexplained variance explained by 1st contrast < 5% is excellent.	Slightly Yes

3.4. Reliability

Reliability refers to the consistency of test results and supports generalizability validity, indicating whether the reliability and precision of an instrument are satisfactory. Three reliability measures were investigated using results from Winsteps: Cronbach's Alpha, Rasch reliability, and Person Separation (Wright, 1996). Cronbach's Alpha, based on the Classical Test Theory framework, was found to be 0.93, indicating good reliability. Rasch reliability, specified as 1 - MSEθ / S^2θ, also demonstrated good reliability at 0.92, surpassing the gold standard of 0.90. Person Separation (G), an alternative to Rasch Person reliability, was calculated as the square root of R / (1 - R), where R represents Rasch Reliability. The Person Separation of CGEI was 3.47, indicating good reliability as it exceeded the standard threshold of 3.

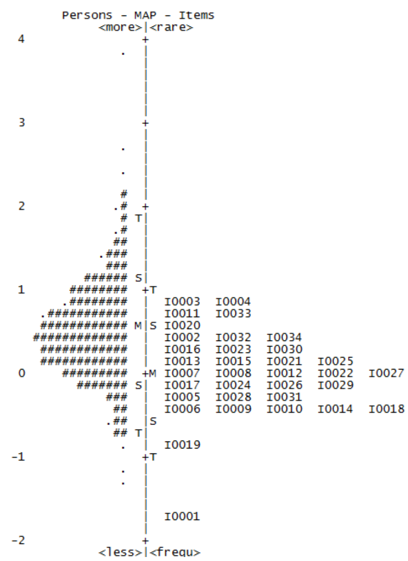
In addition to overall reliability measures, Test Information Function was examined to determine where the instrument exhibited the highest accuracy. Figure 1 revealed that the instrument was most precise for individuals with an ability level around -0.5. Therefore, CGEI is effective in measuring individuals with an average level of computer game engagement, with the highest precision observed for those around -0.5 as their ability levels.

***Figure 1.* Test Information Function**



Furthermore, item hierarchy was assessed through the person-item map depicted in Figure 2 below. The person-item map aids in determining whether item difficulties align with the levels of engagement demonstrated by respondents, thereby supporting the external aspect of validity. The results indicated that some respondents exhibited relatively higher engagement levels ranging from 1 to 2, while most CGEI items were deemed suitable for measuring lower engagement levels ranging from -1 to 1. Thus, CGEI is well-suited for assessing individuals at lower engagement levels, as designed by including lower process levels of Bloom.

Figure 2. **Person-Item Maps**



4. DISCUSSION

In pursuit of a reliable and valid tool to assess the impact of educational games, this study endeavored to develop an instrument with high psychometric quality. Central to this endeavor was the selection of engagement as the primary construct, given its recognized significance in determining student academic success (Archambault et al., 2009; Finn & Rock, 1997; Lamborn et al., 1992; Marks, 2000). The study began by operationally defining game engagement and adopted a comprehensive approach by incorporating three subcomponents: behavioral, cognitive, and affective engagement, in line with recommendations from researchers (Appleton et al., 2006; Fredricks et al., 2004).

Importantly, this study aimed to enhance the quality of instruments by integrating the development and validation processes of the Computer Game Engagement Instrument (CGEI) based on Messick's framework (1995). During the development stage, the CGEI was meticulously crafted to establish Messick's validity by incorporating evidence related to content, substantive, and structural components. Subsequently, at the validation stage, the CGEI underwent evaluation to determine if it provided empirical evidence regarding Messick's validity components. For this evaluation, the study employed the Rasch Partial Credit model for polytomous items. Given the current trend of adopting Messick's validity as an evaluation plan for technology-mediated learning instruments (Filsecker, & Hickey, 2014; Sung, Chang, & Yu, 2011; Webb et al., 2013), the approach undertaken in this study, which utilizes Messick's validity framework as a primary guideline for instrument development and validation, represents a significant contribution to the field. Additionally, by adopting the Rasch model to assess the psychometric properties of engagement instruments, this study extended previous research, which primarily focused on reliability measures in the field of technology-mediated learning (Barkatsas et al., 2009; Junco et al., 2011; Pierce et al., 2007).

During the development stage, the study ensured content validity by following procedures such as developing clearly defined constructs, establishing the rationale for CGEI, and testing specifications based on a comprehensive literature review to ensure content relevance and representativeness. The validation process successfully provided empirical evidence for content validity, with the CGEI demonstrating sufficient technical qualities such as item difficulty and discrimination.

To establish substantive validity, the study provided operational definitions of the three subcomponents of engagement during the development stage, as well as an internal model grounded in a theoretical framework. At the validation stage, the CGEI yielded mixed results in terms of substantive validity. While item hierarchy results demonstrated consistency with expectations, both person fit analysis and rating scale analysis did not fully support substantive validity. Further exploration is warranted to identify potential sources of inconsistency, such as participants' diverse experiences with different types of games.

In terms of structural validity, the study developed an internal model based on a theoretical framework, representing a broad construct of game engagement with three subcomponents. Principal Component Analysis (PCA) results supported the unidimensionality of the CGEI, suggesting structural validity. However, given the contrasting findings in previous research regarding the multidimensionality of engagement constructs (Glanville & Wildhagen, 2007; Wang, Willett, & Eccles, 2011), future studies should explore alternative measurement models before concluding on the unidimensionality of game engagement. The study also recognized the need to expand the range of items to measure higher levels of engagement, as participants tended to exhibit medium to high engagement levels. Despite this, the CGEI demonstrated generalizability and adequate reliability across various measures.

In conclusion, the CGEI presents a reliable and valid measure of engagement in computer games for educational purposes. Practitioners can utilize CGEI as an assessment tool to implement and evaluate educational games, gaining insights into their effectiveness based on learners' engagement levels. Additionally, researchers can conduct rigorous investigations into the impact of computer games on academic engagement and achievement, mediated through game engagement. Furthermore, CGEI has the potential to offer parents guidance in selecting appropriate computer games for their children's educational needs.

References

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science, 12*(5), 353-359.

Annetta, L., Mangrum, J., Holmes, S., Collazo, K., & Cheng, M.-T. (2009). Bridging reality to virtual reality: Investigating gender effect and student engagement on learning through video game play in an elementary school classroom. *International Journal of Science Education, 31*(8), 1091 - 1113.

Archambault, I., Janosz, M., Fallu, J. S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence*, *32*(3), 651-670.

Barkatsas, A. T., Kasimatis, K., & Gialamas, V. (2009). Learning secondary mathematics with technology: Exploring the complex interrelationship between students' attitudes, engagement, gender and achievement. *Computers & Education*, *52*(3), 562-570.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Filsecker, M., & Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. *Computers & Education*, *75*, 136-148.

Fincham, F. R., Hokoda, A., & Sanders, R. (1989). Learned helplessness, test anxiety, and academic achievement: A longitudinal analysis. *Child Development, 60,* 138–145.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117-142.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, *82*(2), 221.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research, 74*(1), 59-109.

Funk, J. B. (2002). Electronic games. In V. Strasburger & B. Wilson (Eds.), Children, adolescents, and the media (pp. 117–144). Thousand Oaks, CA: Sage.

Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement assessing dimensionality and measurement invariance across race and ethnicity. *Educational and Psychological Measurement*, *67*(6), 1019-1041.

Hoffman, B., & Nadelson, L. (2010). Motivational engagement and video gaming: a mixed methods study. *Educational Technology Research and Development, 58*(3), 245-270.

Holtz, P., & Appel, M. (2011). Internet use and video gaming predict problem behavior in early adolescence. *Journal of Adolescence, 34*(1), 49-58.

Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, *27*(2), 119-132.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education, 51*(4)*,* 1609-1620.

Lamborn, S., Newmann, F., & Wehlage, G. (1992). The significance and sources of student engagement. *Student engagement and achievement in American secondary schools*, 11-39.

Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8*(2), 139 – 162.

Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis. *Rasch Measurement Transactions*, *12*(2), 636.

Linacre, J. M. (2005). Winsteps® (Version 3.58.0) [computer software]. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2005). A user's guide to WINSTEPS MINISTEP Rasch-model computer programs*. Chicago: Winsteps.com.*

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153-184.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Meyer, J. P. (2009). *jMetrik version 1* [computer software]. Retrieved from www.ItemAnalysis.com.

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English, 25,* 261–290.

Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1-12.

Perry, J. C., Liu, X., & Pabian, Y. (2010). School engagement as a mediator of academic performance among urban youth: The role of career preparation, parental career support, and teacher support. *The Counseling Psychologist*, *38*(2), 269-295.

Pierce, R., Stacey, K., & Barkatsas, A. (2007). A scale for monitoring students' attitudes to learning mathematics with technology. *Computers & Education*, *48*(2), 285-300.

Prensky, M. (2001). *Digital game-based learning.* New York: McGraw-Hill.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321-333). Berkeley, CA: University of California Press.

Simons-Morton, B., & Chen, R. (2009). Peer and parent influences on school engagement among early adolescents. *Youth & Society*, *41*(1), 3-25.

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1*, 199-218.

Sung, Y. T., Chang, K. E., & Yu, W. C. (2011). Evaluating the reliability and impact of a quality assurance system for E-learning courseware. *Computers & Education*, *57*(2), 1615-1627.

Voelkl, K. E. (1997). Identification with school. *American Journal of Education, 105,* 204–319.

Wang, M. T., Willett, J. B., & Eccles, J. S. (2011). The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology*, *49*(4), 465-480.

Webb, M., Gibson, D., & Forkosh-Baruch, A. (2013). Challenges for information technology supporting educational assessment. *Journal of Computer Assisted Learning*, *29*(5), 451-462.

Willoughby, T. (2008). A short-term longitudinal study of Internet and computer game use by adolescent boys and girls: Prevalence, frequency of use, and psychosocial predictors. *Developmental Psychology, 44*(1), 195-204.

Wright, B. D. (1996). Reliability and separation. *Rasch Measurement Transactions*, *9*(4), 472.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Appendix. Computer Game Engagement Instrument and Items

	Strongly disagree					Stroonly agree	
I followed the computer game rules when I played the game.	1	2	3	4	5	6	N/A
I completed game tutorials or practice games the first time when I played the computer game.	1	2	3	4	5	6	N/A
When I did not understand the rules of the game, I asked questions to a friend who knows well about the game.	1	2	3	4	5	6	N/A
I shared my computer game experiences with other friends through the Internet, verbal, or phone conversation.	1	2	3	4	5	6	N/A
I played computer games until I reached my desired level.	1	2	3	4	5	6	N/A
I paid attention to the game's instructions and directions.	1	2	3	4	5	6	N/A
I did the game activities with a lot of effort.	1	2	3	4	5	6	N/A
I chose to do hard tasks rather than easy tasks while playing the games.	1	2	3	4	5	6	N/A
I tried to finish the game even when it was difficult.	1	2	3	4	5	6	N/A
I attempted to play games thoroughly and well, rather than just trying to get by.	1	2	3	4	5	6	N/A
I talked about the game with friends before or after the game.	1	2	3	4	5	6	N/A
I was willing to work hard to understand tasks that computer games required.							
I was willing to do challenging tasks of the computer game.	1	2	3	4	5	6	N/A
I felt happy when playing the computer game.	1	2	3	4	5	6	N/A
I often lost track of time and place when playing the computer game.	1	2	3	4	5	6	N/A
The computer game was good to improve thinking skills.	1	2	3	4	5	6	N/A
After I played the computer game, I felt I improved my technology skills.	1	2	3	4	5	6	N/A
Learning about the computer game would be helpful for my future job employment in the long run.	1	2	3	4	5	6	N/A
I was willing to do challenging tasks of the computer game.	1	2	3	4	5	6	N/A

I enjoyed most game activities.	1	2	3	4	5	6	N/A
I believe that computer games can be used as an important tool for learning.	1	2	3	4	5	6	N/A
I think highly of my friends who play computer games well.	1	2	3	4	5	6	N/A
I felt bored when playing computer games.	1	2	3	4	5	6	N/A
I got discouraged and stopped trying when I encountered an obstacle in computer games.	1	2	3	4	5	6	N/A
I became easily frustrated during games.	1	2	3	4	5	6	N/A
I was flexible to use different strategies to accomplish difficult tasks of the computer game.	1	2	3	4	5	6	N/A
I made efforts to master high-level game skills.	1	2	3	4	5	6	N/A
I continued to play computer games even when I failed in a certain game.	1	2	3	4	5	6	N/A
I was willing to improve my understanding/skills of the computer game.	1	2	3	4	5	6	N/A
It was easy for me to set up a goal while playing the computer game.	1	2	3	4	5	6	N/A
I organized my time and resources well to accomplish the game goal.	1	2	3	4	5	6	N/A
I had a clear idea of what I was trying to accomplish while I was playing the computer game.	1	2	3	4	5	6	N/A
Before I began to play games, I made sure I knew what I was asked to do.	1	2	3	4	5	6	N/A
I tried to organize an approach in my mind before I actually started games.	1	2	3	4	5	6	N/A
When I played games, I analyzed it to see if there was more than one way to accomplish the game goal.	1	2	3	4	5	6	N/A