
Volume 27, Issue 5, September 2011

ISSN 0747-5632

ELSEVIER

COMPUTERS IN HUMAN BEHAVIOR

Robert D. Tennyson, Editor
Paul Kirschner, Associate Editor
Jen-Her Wu, Associate Editor

Including Special Section
2009 Fifth International Conference on Intelligent Computing (ICIC 2009)

Guest Editor
De-Shuang Huang

Computers in Human Behavior 27 (2011) 1849–1856

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Evaluating learning, design, and engagement in web-based learning tools (WBLTs): The WBLT Evaluation Scale

Robin Kay

University of Ontario, Institute of Technology, Faculty of Education, 11 Simcoe St. North, Oshawa, Ontario, Canada L1H 7L7

ARTICLE INFO

Article history:
Available online 24 May 2011

Keywords:
Evaluate
Learning
Design
Scale
Web-based learning tools
Learning object

ABSTRACT

Web-based learning tools (WBLTs), also known as learning objects, are online, interactive tools that support the learning of specific concepts by enhancing, amplifying, and/or guiding the cognitive processes of learners. Research examining the effectiveness of WBLTs is somewhat limited because sound, reliable, valid evaluation metrics are sparse, particularly in the K-12 environment. The purpose of the following study was to re-examine the Learning Object Evaluation Scale for students (LOES-S), originally developed by Kay and Knaack (2009), to assess three key constructs: learning, design, and engagement. Over 800 middle and secondary schools students participated in high quality, pre-designed lessons intended to accentuate the use of WBLTs. Data collected from the new WBLT Evaluation Scale demonstrated good internal reliability, construct validity, convergent validity and predictive validity.

1. Overview

Web-based learning tools, also known as learning objects, are operationally defined in this study as interactive, online learning tools that support the learning of specific concepts by enhancing, amplifying, and/or guiding the cognitive processes of learners. Until recently (Kay & Knaack, 2009), research on the effectiveness of WBLTs was limited (e.g., Kay & Knaack, 2005; Nurmi & Jaakkola, 2005, 2006a,b; Sosteric & Hesemeier, 2002). Most studies evaluating WBLTs focussed on development and design, but not impact on students in the classroom (e.g., Bradley & Boyle, 2004; Vargo, Nesbit, Belfer, & Archambault, 2003; Williams, 2000). Assessing effectiveness in the classroom, though, is critical if WBLTs are to be considered as effective educational tools.

The majority of WBLTs use and evaluation has occurred in higher education (e.g., Cochrane, 2005; Haughey & Muirhead, 2005; Kay & Knaack, 2005; Krauss & Ally, 2005; Nesbit & Belfer, 2004). However, increased use of WBLTs in middle and secondary schools (e.g., Brush & Saye, 2001; Clarke & Bowe, 2006a, 2006b; Kay & Knaack, 2005; Liu & Bera, 2005; Lopez-Morteo & Lopez, 2007; Nurmi & Jaakkola, 2006a) suggests that there is a need to develop evaluation metrics for the K-12 domain.

In 2009, Kay and Knaack developed the Learning Object Evaluation Scale for Students (LOES-S) to examine the impact of learning objects on students in grades 7–12. The LOES-S showed good reliability and validity. The purpose of the current study was to build on the results reported by Kay and Knaack (2009) by controlling the selection of WBLTs, design of lessons plans, and the quality of learning performance assessment.

2. Literature review

2.1. Definition of web-based learning tools (WBLTs)

Web-based learning tools (WBLTs) have also been referred to "learning objects". However, extensive debate has clouded the definition of learning objects producing a wide range of descriptions including any tool (online or otherwise) that supports learning, online artefacts, interactive, re-usable tools that enhance learning, and entire online courses (Bennett & McGee, 2005; Bradley & Boyle, 2004; Cochrane, 2005; Koppi, Bogle, & Bogle, 2005; McGreal, 2004; Parrish, 2004; Siqueira, Melo, & Braz, 2004; Wiley et al., 2004). To date, consensus on an acceptable definition has not been achieved.

Rather than weighing into this unresolved debate, it was decided that the term "web-based learning tool" would be used to eliminate any possible confusion. Key factors of WBLTs include interactivity, accessibility, a specific conceptual focus, meaningful scaffolding, and learning. WBLTs are operationally defined in this study as "online, interactive tools that support the learning of specific concepts by enhancing, amplifying, and/or guiding the cognitive processes of learners". Specific WBLTs used in this study allowed students to experiment, manipulate variables, apply concepts, or answer questions based on formal presentation of material targeting a relatively narrow concept. To view specific examples of the WBLTs used by teachers in this study, see Kay (2011).

E-mail address: robin.kay@uoit.ca

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.04.007

2.2. Previous evaluation of WBLTs

Kay and Knaack (2009) noted that research on the evaluation of WBLTs was limited in two fundamental ways. First, most studies examining the quality of WBLTs lacked a coherent assessment structure. Second, the methodology used to evaluate WBLTs has been somewhat restricted in breadth and scope. Each of these limitations will be discussed in more detail.

2.3. Underlying structure of previous WBLT Evaluation metrics

Kay and Knaack (2009) noted that the majority of evaluation methods and tools used to measure WBLTs have limited structure and organization (e.g., Buzzetto-More & Pinhey, 2006; Gadanidis, Sedig, & Liang, 2004; Koohang & Du Plessis, 2004; McGreal et al., 2004; Schoner, Buzza, Harrigan, & Strampel, 2005). Instead, a wide range of isolated features have been examined including user control, interactivity, layout, personalization, quality of graphics, difficulty level, theme, aesthetics, feedback, range of multimedia used, ease of use, clarity of instructions, navigation, accuracy, and quality of content (Kay & Knaack, 2008). After conducting several extensive reviews of the WBLT Evaluation literature, three key constructs for assessing WBLTs were identified: learning, design, and engagement (Kay & Knaack, 2005, 2007, 2008, 2009).

With respect to the learning construct, the following features were thought to be critical including *interactivity* (Lim, Lee, & Richards, 2006; Ohl, 2001; Oliver & McLoughlin, 1999; Van Merriënboer & Ayres, 2005), *good quality feedback* (Brown & Voltz, 2005; Kramarski & Zeichner, 2001; Nielson, 2003; Reimer & Moyer, 2005), *visual supports* (Gadanidis et al., 2004; Haughey & Muirhead, 2005; Nesbit & Belfer, 2004; Oliver & McLoughlin, 1999) and *whether new concepts were learned* (Kay & Knaack, 2005, 2007).

Regarding the design construct, previous studies indicated that the following components were important: *clarity of instructions and help features* (Acovelli & Gamble, 1997; Jones, Farquhar, & Surry, 1995; Kennedy & McNaught, 1997), *ease of use* (Haughey & Muirhead, 2005; Lin & Gregor, 2006; Macdonald et al., 2005; Schell & Burns, 2002; Schoner et al., 2005), and *overall organization and layout* (Calvi, 1997; Del Moral & Cernea, 2005; Madhumita, 1995; Mayer & Moreno, 2002).

Finally, research on the engagement construct suggested that *overall theme* (Haughey & Muirhead, 2005; Jonassen and Churchhill, 2004; Kay & Knaack, 2005; Lin & Gregor, 2006; Macdonald et al., 2005; Reimer & Moyer, 2005; Van Zele, Vandaele, Botteldooren, & Lenaerts, 2003), *multimedia used* (Brown & Voltz, 2005; Nesbit & Belfer, 2004; Oliver & McLoughlin, 1999), and *willingness to use WBLTs again* (Kay and Knaack, 2005, 2007) were central qualities to assess.

Kay and Knaack (2009) developed the Learning Object Evaluation Scale based on the learning, design, and engagement constructs. Tested on over 1100 middle and secondary school students, the LOES-S showed acceptable internal reliability, face validity, construct validity, convergent validity and predictive validity (Kay & Knaack, 2009). It is reasonable to conclude that the structure of the LOES-S is a promising first step in evaluating the effectiveness and quality of WBLTs. One key goal of the current study is to revise and re-test the three prong structure originally proposed by Kay and Knaack (2009).

2.4. Methodology and WBLT Evaluation

In an effort to improve the quality of WBLT Evaluation research, Kay and Knaack (2009) responded to six key patterns in methodology. These included (a) excessive energy directed toward assessing the development of WBLTs as opposed to impact on student learning (e.g., Bradley & Boyle, 2004; Cochrane, 2005), (b) the predominance of qualitative data (e.g., Kenny, Andrews, Vignola, Schilz, & Covert, 1999; Lin & Gregor, 2006), (c) focussing on a narrow range of WBLTs (e.g., Bradley & Boyle, 2004; Krauss & Ally, 2005; MacDonald et al., 2005), (d) small and poorly described sample populations (e.g., Cochrane, 2005; Krauss & Ally, 2005; MacDonald et al., 2005; Van Zele et al., 2003), (e) a noticeable absence of reliability, validity and statistical data (e.g., Howard-Rose & Harrigan, 2003; Lopez-Morteo & Lopez, 2007; Schoner et al., 2005; Vacik, Wolfslehner, Spork, & Kortschak, 2006; Vargo et al., 2003), and (f) a recent shift toward including learning performance data (e.g., Bradley & Boyle, 2004; Docherty, Hoy, Topp, & Trinder, 2005; MacDonald et al., 2005; Nurmi & Jaakkola, 2006a).

Kay and Knaack (2009) addressed many of the methodological concerns when they developed and analysed the LOES-S. A large, diverse sample of middle and secondary school students used a wide range of WBLTs. Reliable and valid quantitative data were collected as well as learning performance scores.

However, there were three potentially noteworthy problems in Kay and Knaack's approach to evaluating WBLTs. First, the choice of WBLTs was essentially random. Teachers searched for and selected a wide range of WBLTs without structure or substantial guidance. Therefore, the WBLTs varied considerably and included the simple presentation of facts, short 5 min games, detailed websites on a specific topic as well as more traditional WBLTs that were interactive with visual supports. It is unclear whether some of the tools used matched Kay and Knaack's (2009) proposed operational definition of WBLTs. Therefore the type of WBLT used may have been a confounding variable. In order to control for this factor, in the current study a large database of WBLTs was pre-selected by trained teachers based on the Kay and Knaack's (2008) multi-component model for evaluating WBLTs.

The second issue with Kay and Knaack's study was the wide range of teaching strategies used with WBLTs. Although teachers in the study received a day of training, the range of WBLT activities used in their classrooms was considerable. Teachers used WBLTs as a hook, a review or previous concepts learned, an exploration tool, consolidations of concepts taught using another teaching method, or a fun activity at the end of a lesson. In addition, the time spent using a WBLT ranged from 5 min to an entire 60 min lesson. This variability makes it difficult to decipher what students are evaluating when they use WBLTs. In order to control the impact of teaching strategy, in this study a set of pre-designed lesson plans were created by experienced teachers based on previous research looking at effective strategies for using WBLTs (Kay, Knaack, & Muirhead, 2009).

The final concern with Kay and Knaack's methodology was the variability in learning performance measures. Teachers who used the WBLTs created their own pre- and post-tests, therefore the depth and scope of assessment was non-standardized. In addition, learning performance was assessed with a single score and did not account for type of knowledge learned (e.g., remembering vs. applying vs. analyzing). Two procedures were followed to address learning performance issues. First, an enhanced measure of student performance was created by trained teachers for each WBLT used in the study. Second, four knowledge categories were targeted based on the revised Bloom taxonomy (Anderson & Krathwohl, 2001).

2.5. Purpose

The purpose of this study was to revise and re-evaluate Kay and Knaack's (2009) student-focused, learning-based approach for assessing WBLTs. The original structure of Kay and Knaack's assessment tool was unchanged and included learning, design, and engagement constructs. However, three major revisions were made on the original study to control of potential confounding influences including pre-selecting WBLTs, pre-design of lesson plans, and developing customized measures of learning performance based on the revised Bloom taxonomy (Anderson & Krathwohl, 2001) and the nature of the WBLT used.

3. Method

3.1. Sample

3.1.1. Students

This sample consisted of 834 students (392 males, 441 females, one missing data), 11–17 years of age ($M = 13.3$, $SD = 0.97$), from middle ($n = 444$) and secondary ($n = 390$) schools. A majority of students were enrolled in grades seven ($n = 229$), eight ($n = 215$), and nine ($n = 340$). Over 75% ($n = 628$) of the students claimed their average mark was 70% or more in the subject area where the WBLT was used. In addition, over three quarters of students agreed or strongly agreed that that they were good at working with computers. The sample population was gleaned from 33 different secondary school classes located within a sub-urban region of almost 600,000 people.

3.1.2. Teachers

This sample consisted of 28 teachers (eight males, 20 females). Nine teachers taught grade 7, nine teachers taught grade 8, seven teachers taught grade 9, and three teachers taught grade 10. Subjects taught were mathematics ($n = 15$) and science ($n = 13$). Class size ranged from 9 to 28 with a mean of 18 students ($SD = 5.4$). Teaching experience varied from 0.5 to 23 years with a mean of 7.1 ($SD = 6.7$). Over 80% ($n = 23$) of the teachers agreed that they (a) were good at working with computers and (b) liked working with computers at school.

3.1.3. WBLTs and lesson plans

Four teachers (not participants in the study) were trained for 2 days on how to select WBLTs for the classroom and develop effective lesson plans. The criteria for selecting WBLTs was based on Kay and Knaack's (2008) multi-component model for assessing WBLTs. The lesson plan design evolved from the results of a previous research study by Kay, Knaack, and Muirhead (2009). The key components of these lesson plans included a guiding set of questions, a structured well-organized plan for using the WBLTs, and time to consolidate concepts learned. Over a period of 2 months, a database of 122 lesson plans and WBLTs was created (78 for mathematics and 44 for science). A total of 22 unique WBLTs were selected by the classroom teachers in this study from the WBLT database. A wide variety of WBLTs were used involving experimentation, virtual manipulatives, task-based applications, and formal presentation concepts followed by a question and answer assessment. See Appendix A (Kay, 2011) for a complete list of all WBLTs and associated lesson plans used.

3.2. Procedure

Teachers from two boards of education were emailed by an educational coordinator and informed of the WBLT study. Participation was voluntary and participants could withdraw at any time. Each teacher received a full day of training on using and implementing the pre-designed WBLT lesson plans. They were then asked to use at least one WBLT in their classroom. Email support was available for duration of the study. All students in a given teacher's class used the WBLT that the teacher selected, however, only those students with signed parental permission forms were permitted to fill in an anonymous, online survey (Appendix A). In addition, students completed pre- and post-tests based on the content of the WBLT. These tests were pre-designed by the authors of the lesson plans to match the teaching goals of the WBLT.

3.3. Data sources

3.3.1. Student survey

After using a WBLT, students completed the WBLT Evaluation Scale (see Appendix A) to determine their perception of (a) how much they learned (learning construct), (b) the design of the WBLT (design construct), and (c) how much they were engaged with the WBLT (engagement construct). Descriptive statistics for the WBLT Evaluation Scale are presented in Table 1.

3.3.2. Student performance

Students completed a pre- and post-test based on the content of the WBLT used in class. These tests were included with all pre-designed lesson plans to match the learning goals of the WBLT. The difference between pre- and post-test scores was used to determine changes in student performance on four possible knowledge categories: remembering, understanding, application, and analysis. These categories were derived from the revised Bloom's Taxonomy (Anderson & Krathwohl, 2001). The number of Bloom's knowledge categories assessed varied according to the learning goals and type of the specific WBLT used.

3.3.3. Teacher survey

After using a WBLT, each teacher completed the WBLT Evaluation Scale for Teachers to determine their perception of (a) how much their students learned (learning construct), (b) the design of the WBLT (design construct), and (c) how much their students were engaged with the WBLT (engagement construct). Data from this scale showed moderate to high reliability in this study (0.93 for learning construct, 0.78 for WBLT design construct, and 0.79 for engagement construct). In addition, Kay, Knaack, and Petrarca (2009) reported good construct validity for this scale using a principal components factor analysis.

3.4. Data analysis

Seven analyses were run to assess the reliability and validity of the WBLT Evaluation Scale. These included:

Table 1
Description of WBLT Evaluation Scale ($n = 823$).

Construct	No. items	Possible range	Actual range observed	Internal reliability
Learning	5	5–35	5–35	$r = 0.93$
Design	4	4–28	4–28	$r = 0.87$
Engagement	4	4–28	4–28	$r = 0.92$

Table 2
Varimax rotated factor loadings on WBLT Evaluation Scale.

Scale item	Factor 1	Factor 2	Factor 3
S-learning 1 – learn	.759		
S-learning 2 – feedback	.772		
S-learning 3 – graphics	.723		
S-learning 4 – new concept	.798		
S-learning 5 – overall	.740		
S-design 6 – help		.596	
S-design 7 – instructions		.823	
S-design 8 – easy to use		.854	
S-design 9 – organized		.744	
S-engagement 10 – theme			.737
S-engagement 11 – engaging			.783
S-engagement 12 – fun			.821
S-engagement 12 – use again			.798
Factor	Eigen value	PCT of VAR	Cum PCT
1	8.16	62.8	62.8
2	1.10	8.5	71.3
3	0.87	6.7	78.0

(1) internal reliability estimates (reliability);
(2) a principal component factor analysis for WBLT Evaluation Scale (construct validity);
(3) correlations among WBLT Evaluation Scale constructs (construct validity);
(4) correlations among WBLT Evaluation Scale constructs and teacher ratings (convergent validity);
(5) correlation between WBLT Evaluate Scale constructs and computer comfort level (convergent validity);
(6) correlation between WBLT Evaluate Scale constructs and subject area comfort level (convergent validity);
(7) correlation between WBLT Evaluate Scale constructs and learning performance measures (predictive validity);

4. Results

4.1. Internal reliability

The internal reliability estimates for the WBLT Evaluation Scale constructs based on Cronbach's α were 0.93 (learning), 0.87 (design), and 0.92 (engagement) – see Table 1. These values are acceptable for measures used in the social sciences (Kline, 1999; Nunnally, 1978).

4.2. Construct validity

4.2.1. Principal component analysis

A principal components analysis was done to explore whether the WBLT Evaluation Scale constructs (learning, design, and engagement) were three distinct factors. Since all communalities were above 0.4 (Stevens, 1992), the principal component analysis was deemed an appropriate exploratory method (Guadagnoli & Velicer, 1988). Orthogonal (varimax) and oblique (direct oblimin) rotations were used, given that the correlation among potential constructs was unknown. These rotational methods produced identical factor combinations, so the results from the varimax rotation (using Kaiser normalization) are presented because they simplify the interpretation of the data (Field, 2005). The Kaiser–Meyer–Olkin measure of sampling adequacy (0.945) and Bartlett's test of sphericity ($p < .001$) indicated that the sample size was acceptable.

The principal components analysis was set to extract three factors (Table 2). The resulting rotation corresponded with the proposed WBLT Evaluation constructs. The structure was consistent with previous research (Kay & Knaack, 2005, 2007, 2009) and the projected grouping of scale items listed in Appendix A.

4.2.2. Correlations among LOES-S constructs

The correlations between the learning construct and the design ($r = .71$, $p < .001$) and engagement ($r = .76$, $p < .001$) constructs were significant, as was the correlation between the engagement and design construct ($r = .65$, $p < .001$). Shared variances, ranging from 42% to 56% were small enough to support the assumption that each construct measured was distinct.

4.3. Convergent validity

4.3.1. Correlation between WBLT Evaluation Scale and teacher ratings

Mean student perceptions of learning were significantly correlated with teacher ratings of learning and design. but not engagement. Mean student ratings of WBLT design were significantly correlated with all three teacher rated constructs. Finally, student assessment of WBLT engagement was significantly correlated with teacher ratings of learning and design. but not engagement.

Overall, correlations ranging from 0.36 to 0.65 showed a moderate degree of consistency between student and teacher evaluations WBLTs (Table 3).

Table 3
Correlations among WBLT Evaluation constructs and teacher ratings ($n = 33$).

	Learning (students) r	Design (students) r	Engagement (students) r
Learning (teachers)	0.56***	0.52***	0.50***
Design (teachers)	0.46**	0.40*	0.65***
Engagement (teachers)	0.32	0.36*	0.25

* $p < .05$ (two-tailed).
** $p < .01$ (two-tailed).
*** $p < .005$ (two-tailed).

4.3.2. Correlation between WBLT Evaluation Scale and computer comfort level

Computer comfort level based on a three-item scale (Kay & Knaack, 2007, 2009) was significantly correlated with the learning ($r = .29$, $p < .001$), design ($r = .28$, $p < .001$) and engagement constructs ($r = .30$, $p < .001$). Students who are more comfortable using computers rated the learning, design, and engagement of a WBLT higher.

4.3.3. Correlation between WBLT Evaluation Scale and subject area comfort level

Subject area comfort level was assessed using two item scale focussing on a student's perception of how good they were at and how much they liked the subject area covered by WBLT they used. The internally reliability for the scale was 0.77. Subject comfort area was significantly correlated with the learning ($r = .29$, $p < .001$), design ($r = .31$, $p < .001$) and engagement constructs ($r = .28$, $p < .001$). Students who are more comfortable with the subject area covered by a WBLT rated the learning, design, and engagement of a WBLT higher.

4.4. Predictive validity

4.4.1. Correlation between WBLT Evaluation Scale and learning performance

Four categories of learning performance were assessed (remembering, understanding, application, and analysis). Student perceptions of learning, design, and engagement of WBLTs were significantly correlated with positive gains in application and analysis knowledge areas, but not remembering or understanding. In other words, higher scores on student perceptions of learning, WBLT design, and engagement were associated with higher scores in learning performance in two of the four learning performance categories assessed, although the magnitude of the correlation coefficients was relatively small (Table 4).

5. Discussion

The purpose of this study was to revise and re-evaluate the LOES-S tested by Kay and Knaack in 2009. The original model for the LOES-S was student-focused and based on three prominent themes that appeared in previous WBLT Evaluation research: learning, design, and engagement. This model was also used for the current study. Three revisions were made to the original study in order to control for confounding variables and included pre-selected WBLTs, pre-designed lesson plans, and enhanced measures of learning performance that incorporated four knowledge areas (remembering, understanding, application, analysis) from the revised Bloom taxonomy (Anderson & Krathwohl, 2001).

5.1. Sample population and variety of WBLTs

The sizeable population in this study was selected primarily from grade 7–9 math and science classrooms ($n = 834$) located in a relatively large sub-urban area. The size and diversity of

Table 4
Correlations among WBLT Evaluation constructs and learning performance measures.

	Learning (students) r	Design (students) r	Engagement (students) r
Remembering ($n = 418$)	0.01	−0.08	−0.04
Understanding ($n = 253$)	0.11	0.04	0.01
Application ($n = 418$)	0.16**	0.12*	0.16**
Analysis ($n = 87$)	0.37***	0.30**	0.31**

* $p < .05$ (two-tailed).
** $p < .005$ (two-tailed).
*** $p < .001$ (two-tailed).

the current sample helps build on previous smaller scale research and provides greater confidence in the data collected and conclusions drawn. In addition, the wide range of WBLTs used suggests that the usefulness of the WBLT Evaluation Scale extends beyond a narrow set of WBLTs. Based on the data from Kay and Knaack (2009) and the present study, it is reasonable to assume that the WBLT Evaluation Scale is a tool that can be used to evaluate a wide variety of WBLTs in both middle and secondary school environments.

5.2. Reliability

The internal reliability estimates (0.87–0.93) for the three constructs in the WBLT Evaluation were high (Kline, 1999; Nunnally, 1978). The values were higher than the original LOES-S, a result that may reflect the care and attention that went into selecting WBLTs and creating custom-made lesson plans. According to Kay and Knaack (2009), less than one quarter of WBLT Evaluation studies offer some form of reliability statistics. However, reliability is a fundamental element of any high quality evaluation tool. Given that internal reliability was good for both the LOES-S and the WBLT Evaluation Scale, it is fair to conclude that the three constructs (learning, design, and engagement) demonstrate internally reliability.

5.3. Validity

Kay and Knaack (2009) noted a small fraction of WBLT Evaluation studies assess validity (Kay & Knaack, 2007; Nurmi & Jaakkola, 2006a). Three types of validity were considered in this paper: construct, convergent, and predictive.

5.3.1. Construct validity

The factor analysis revealed three distinct WBLT constructs that were consistent with the theoretical framework proposed by Kay and Knaack (2009). In the original study, some overlap existed among scale items were observed, however, no overlap appeared within WBLT Evaluation Scale constructs in this study. Controlling for type of WBLTs and quality of lesson plans might have helped improved construct validity.

Relatively high correlations (0.63–0.74) demonstrated that learning, design, and engagement constructs were inter-related. However, shared variances of 42–56% indicated that the learning, design, and engagement constructs were also distinct.

It is important to recognize that while it is statistically advantageous to identify discrete constructs in order to discover potential strengths and weaknesses of WBLTs, it is highly likely that these constructs interact and mutually influence each other when actual learning occurs. For example, a WBLT that is not effective at promoting learning could have a negative impact on students perceptions of engagement. On the other hand, a WBLT that is poorly designed and difficult to use may frustrate students to the point of impeding learning. Finally, a highly engaging WBLT may focus students attention thereby leading to more effective learning. It is reasonable to conclude that the constructs do share explanatory variance with respect to evaluating WBLTs, but that they also represent three unique features of WBLTs: learning, design, and engagement.

5.3.2. Convergent validity

Convergent validity was analysed using three tests. First, correlations between student estimates of learning and design were significantly correlated with teacher estimates of these same constructs. However, student perceptions of engagement did not significantly correlate with teacher's perceptions of engagement. This result has been reported in several studies (Kay & Knaack, 2008, 2009). All correlations were modest with a shared variance ranging from 13% to 40%. These results might be anticipated given that teachers and students may have different perspectives on what constitutes learning, design, and engagement. Therefore, while student and teacher constructs do converge, the modest correlations underline one of the main premises of this paper, namely the need for obtaining student input.

The second test of convergent validity looked at correlations among the three WBLT Evaluation Scale constructs and student computer comfort level. It was assumed that student comfort level with computers would influence ratings of learning, design, and engagement of WBLTs. This assumption was supported by the modest (0.28–0.30), but significant correlation estimates observed. This result was also confirmed by Kay and Knaack (2009). The relatively low correlations among computer comfort level and the three WBLT Evaluation Scale constructs may be partially explained by how easy it is to use most WBLTs. Students who are uncomfortable with computers may experience minor, but not excessive challenges when interacting with WBLTs.

The third test of convergent validity examined correlations among WBLT Evaluation Scale constructs and student comfort level with the subject areas covered by the WBLT being used. Kay and Knaack (2009) suggested that subject area comfort level be tested in future research. They hypothesized that the more comfortable a student was with a specific subject area, the more likely he/she is to have a accepting attitude about using a new WBLT. On the other hand, if a student was not comfortable with a subject area, he/she might be more resistant to the use of a new tool. Significant positive correlations among student perceptions of WBLT learning, design, and engagement and subject comfort level confirmed Kay and Knaack's (2009) hypotheses.

5.3.3. Predictive validity

It is reasonable to predict that WBLTs that are rated highly in terms of learning, design, and/or engagement would be positively correlated with higher learning performance. In other words, if a student perceives a WBLT as a well-designed, engaging learning tool we would expect him/her to perform better in a pre-post test situation. This prediction was confirmed by the original LOES-S study of Kay and Knaack (2009). However, the quality and scope of the learning performance measures were limited.

In the current study, four knowledge categories based on the revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) were assessed using pre- and post-tests that were custom designed to address the specific learning goals of each WBLT. Student perceptions of learning, design, and engagement were significantly correlated with increases in knowledge categories focussing on application and analysis, but not remembering or understanding.

A possible explanation for this result is that all WBLTs were preselected based on Kay and Knaack's (2008) multi-component model for evaluating WBLTs. One could safely assume that these WBLTs were of good quality. Therefore, students may have been able to perform well when simpler knowledge areas were addressed like remembering and understanding, regardless of there perceptions of learning quality, design, and engagement. On the other hand, WBLTs may need to be better deigned and more engaging for students to learn higher level concepts involving application and analysis. In other words, student perceptions of WBLT quality are more relevant when the concepts being learned are more advanced. More research is needed to conform this speculation, although it is clear that expanding the range of learning performance measure presents a more complicated pattern of results than Kay and Knaack reported (2009).

5.4. Implications for education

The main purpose for this paper was to revise and re-retest a reliable, valid student-based evaluation tool for assessing WBLTs. The immediate benefit is that future researchers will have an effective metric for assessing the impact of WBLTs from the perspective of middle and secondary school students. However, there are several implications for education. First, it is worthwhile gathering student input before, during, and after using WBLTs. While teacher and student assessment of learning benefits, design, and engagement are consistent with each other, they only share a common variance of 20–40%. It is through student feedback that these tools and associated teaching strategies can be improved.

Second, the WBLT Evaluation Scale offers clear suggestions on key components to focus on when choosing a WBLT. Learning features such as interactivity, clear feedback, and graphics or animations that support learning are desirable, as are design qualities such as effective help, clear instructions, transparency of use and organization. However, the results also suggest that it may be more difficult for a teacher to fully understand what engages a student.

Third, it might be particularly important to be selective when choosing WBLTs that address higher level concepts such application and analysis. Student perceptions of learning, design and engagement appear to be more relevant when more difficult concepts are being learned. Finally, it is worthwhile to note the low significant correlations among student evaluations of learning, design, and engagement and learning performance. No technology will transform the learning process. WBLTs are tools used in a highly integrated classroom environment where decisions on how to use these tools may have considerably more import than the actual tools themselves (Kay, Knaack, & Muirhead, 2009).

5.5. Caveats and future research

Based on the research of Kay and Knaack (2008, 2009), the reliability and validity of WBLT Evaluation Scale was carefully assessed using large sample and a wide range of WBLTs. In addition, WBLTs were systematically selected ahead of time, custom lesson plans were constructed, and enhanced learning performance measures were employed. Nonetheless, there are several caveats that could be addressed in future research.

First, the subject areas covered were mathematics and science. It is conceivable that WBLTs focussing on different subject areas might yield different results. Second, this is the first time that a multi-dimensional measure of learning performance based on Bloom's revised taxonomy has been used. Further research is needed to confirm the results observed. Finally, WBLTs were used one or two times at most in a classroom. Long term use of WBLTs needs to be examined to determine whether student perceptions are consistent and stable.

5.6. Summary

The purpose of the current study was to replicate Kay and Knaack's (2009) original LOES-Scale for evaluating WBLTs. Prior to Kay and Knaack's study, limited research in K-12 classrooms had been done using well-structured, coherent, reliable, and valid measures. Based on Kay and Knaack's (2009) three-pronged structure examining learning, design and engagement, the WBLT Evaluation Scale was developed and implemented with three revisions that were designed to reduce the influence potential extraneous variables (pre-selected WBLTs, pre-designed lesson plans, expanded measure of learning performance). The results revealed that the constructs for WBLT Evaluation Scale were internally reliable and demonstrated good construct, convergent, and predictive validity. These results are consistent with these observed by Kay and Knaack (2009) and suggest that the Web-Based Evaluation Scale, after being tested on almost 2000 middle and education students, is a reliable, valid metric that can be used to assess the quality of WBLTs. Evidence was also presented that student perceptions of learning, design, and engagement may matter more when a WBLT addresses higher level concepts.

Appendix A. WBLT Evaluation Scale

	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
Learning					
1. Working with the learning object helped me learn	1	2	3	4	5
2. The feedback from the learning object helped me learn	1	2	3	4	5
3. The graphics and animations from the learning object helped me learn	1	2	3	4	5
4. The learning object helped teach me a new concept	1	2	3	4	5
5. Overall, the learning object helped me learn	1	2	3	4	5
Design					
6. The help features in the learning object were useful	1	2	3	4	5
7. The instructions in the learning object were easy to follow	1	2	3	4	5
8. The learning object was easy to use	1	2	3	4	5
9. The learning object was well organized	1	2	3	4	5
Engagement	1	2	3	4	5
10. I liked the overall theme of the learning object	1	2	3	4	5
11. I found the learning object engaging	1	2	3	4	5
12. The learning object made learning fun	1	2	3	4	5
13. I would like to use the learning object again	1	2	3	4	5

References

Acovelli, M., & Gamble, M. (1997). A coaching agent for learners using multimedia simulations. *Educational Technology, 37*(2), 44–49.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York: Longman.

Bennett, K., & McGee, P. (2005). Transformative power of the learning object debate. *Open Learning, 20*(1), 15–30. doi:10.1080/0268051042000322078.

Bradley, C., & Boyle, T. (2004). The design, development, and use of multimedia learning objects. *Journal of Educational Multimedia and Hypermedia, 13*(4), 371–389. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/18905>.

Brown, A. R., & Voltz, B. D. (2005). Elements of effective e-learning design. *The International Review of Research in Open and Distance Learning, 6*(1). <http://www.irrodl.org/index.php/irrodl/article/view/217/300>.

Brush, T., & Saye, J. (2001). The use of embedded scaffolds with hypermedia-supported student-centered learning. *Journal of Educational Multimedia and Hypermedia, 10*(4), 333–356. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/8439>.

Buzzetto-More, N. A., & Pinhey, K. (2006). Guidelines and standards for the development of fully online learning objects. *Interdisciplinary Journal of Knowledge and Learning Objects, 2*, 96–104. <http://www.ijello.org/Volume2/v2p095-104Buzzetto.pdf>.

Calvi, L. (1997). Navigation and disorientation: A case study. *Journal of Educational Multimedia and Hypermedia, 6*(3/4), 305–320.

Clarke, O., & Bowe, L. (2006a). *The learning federation and the Victorian department of education and training trial of online curriculum content with Indigenous students* (pp. 1–14). <http://www.sofweb.vic.edu.au/edulibrary/public/teachlearn/ict/TLF_DETVIC_indig_trial_mar06.pdf>.

Clarke, O., & Bowe, L. (2006b). *The learning federation and the Victorian department of education and training trial of online curriculum content with ESL students* (pp. 1–16). <http://www.thelearningfederation.edu.au/verve/_resources/report_esl_final.pdf>.

Cochrane, T. (2005). Interactive QuickTime: Developing and evaluating multimedia learning objects to enhance both face-to-face and distance e-learning environments. *Interdisciplinary Journal of Knowledge and Learning Objects, 1*, 33–54. <http://ijklo.org/Volume1/v1p033-054Cochrane.pdf>.

Del Moral, E. & Cernea, D. A. (2005). Design and evaluate learning objects in the new framework of the semantic web. In A. Mendez-Vila, B. Gonzalez-Pereira, J. Mesa Gonzalez, & J.A. Mesa Gonsalez (Eds.), *Recent research developments in learning technologies* (pp. 1–5). <http://www.formatex.org/micte2005/357.pdf>.

Docherty, C., Hoy, D., Topp, H., & Trinder, K. (2005). E-learning techniques supporting problem based learning in clinical simulation. *International Journal of Medical Informatics, 74*(7–8), 527–533. doi:10.1016/j.ijmedinf.2005.03.00.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Gadanidis, G., Sedig, K., & Liang, H. (2004). Designing online mathematical investigation. *Journal of Computers in Mathematics and Science Teaching, 23*(3), 275–298. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/4731>.

Guadagnoli, E., & Velicer, W. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265–275. doi:10.1037/0033-2909.103.2.265.

Haughey, M., & Muirhead, B. (2005). Evaluating learning objects for schools. *Australasian Journal of Educational Technology, 21*(4), 470–490. <http://www.ascilite.org.au/ajet/ajet21/haughey.html>.

Howard-Rose, D., & Harrigan, K. (2003) *CLOE learning impact studies lite: Evaluating learning objects in nine Ontario University courses*. <http://cloe.on.ca/documents/merlotconference10.doc>.

Jonassen, D., & Churchhill, D. (2004). Is there a learning orientation in learning objects? *International Journal on E-Learning, 3*(2), 32–41. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/12813>.

Jones, M. G., Farquhar, J. D., & Surry, D. W. (1995). Using metacognitive theories to design user interfaces for computer-based learning. *Educational Technology, 35*(4), 12–22.

Kay, R. H. (2011). *Appendix A – List of WBLTs used in secondary school study*. <http://tiny.cc/wblt_appendix_a>.

Kay, R. H., Knaack, L., & Petrarca, D. (2009). Exploring teacher perceptions of Web-based learning tools. *5*, 27–50. <http://www.ijello.org/Volume5/IJELLOv5p027-050Kay649.pdf>.

Kay, R., & Knaack, L. (2005). Developing learning objects for secondary school students: A multi-component model. *Interdisciplinary Journal of E-Learning and Learning Objects, 1*, 229–254. <http://www.ijello.org/Volume1/v1p229-254Kay_Knaack.pdf>.

Kay, R. H., & Knaack, L. (2007). Evaluating the learning in learning objects. *Open Learning, 22*(1), 5–28. doi:10.1080/02680510601100135.

Kay, R. H., & Knaack, L. (2008). A multi-component model for assessing learning objects: The learning object evaluation metric (LOEM). *Australasian Journal of Educational Technology, 24*(5), 574–591. <http://www.ascilite.org.au/ajet/ajet24/kay.pdf>.

Kay, R. H., & Knaack, L. (2009). Assessing learning, quality and engagement in learning objects: The learning object evaluation scale for students (LOES-S). *Education Technology Research and Development, 57*(2), 147–168. doi:10.1007/s11423-008-9094-5.

Kay, R. H., Knaack, L., & Muirhead, B. (2009). A formative analysis of instructional strategies for using learning objects. *Journal of Interactive Learning Research, 20*(3), 295–315. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/26233>.

Kennedy, D. M., & McNaught, C. (1997). Design elements for interactive multimedia. *Australian Journal of Educational Technology, 13*(1), 1–22. <http://www.ascilite.org.au/ajet/ajet13/kennedy.html>.

Kenny, R. F., Andrews, B. W., Vignola, M. V., Schilz, M. A., & Covert, J. (1999). Towards guidelines for the design of interactive multimedia instruction: Fostering the reflective decision-making of preservice teachers. *Journal of Technology and Teacher Education, 7*(1), 13–31. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/8955>.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.

Koohang, A., & Du Plessis, J. (2004). Architecting usability properties in the e-learning instructional design process. *International Journal on E-Learning, 3*(3), 38–44. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/5102>.

Koppi, T., Bogle, L., & Bogle, M. (2005). Learning objects, repositories, sharing and reusability. *Open Learning, 20*(1), 83–91. doi:10.1080/0268051042000322113.

Kramarski, B., & Zeichner, O. (2001). Using technology to enhance mathematical reasoning: Effects of feedback and self-regulation learning. *Education Media International, 38*(2/3), 77–82. doi:10.1080/09523980110041458.

Krauss, F., & Ally, M. (2005). A study of the design and evaluation of a learning object and implications for content development. *Interdisciplinary Journal of Knowledge and Learning Objects, 1*, 1–22. <http://ijklo.org/Volume1/v1p001-022Krauss.pdf>.

Lim, C. P., Lee, S. L., & Richards, C. (2006). Developing interactive learning objects for a computing mathematics models. *International Journal on E-Learning, 5*(2), 221–244. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/5543>.

Lin, A., & Gregor, S. (2006). Designing websites for learning and enjoyment: A study of museum experiences. *International Review of Research in Open and Distance Learning, 7*(3), 1–21. <http://www.irrodl.org/index.php/irrodl/article/view/364/739>.

Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology, Research and Development, 53*(1), 5–21. doi:10.1007/BF02504854.

Lopez-Morteo, G., & Lopez, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers and Education, 48*(4), 618–641. doi:10.1016/j.compedu.2005.04.014.

MacDonald, C. J., Stodel, E., Thompson, T. L., Muirhead, B., Hinton, C., Carson, B., et al. (2005). Addressing the eLearning contradiction: A collaborative approach for developing a conceptual framework learning object. *Interdisciplinary Journal of Knowledge and Learning Objects, 1*, 79–98. <http://ijklo.org/Volume1/v1p079-098McDonald.pdf>.

Madhumita, K. K. L. (1995). Twenty-one guidelines for effective instructional design. *Educational Technology, 35*(3), 58–61.

Mayer, R., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.

McGreal, R. (2004). Learning objects: A practical definition. *International Journal of Instructional Technology and Distance Learning, 1*(9). <http://www.itdl.org/Journal/Sep_04/article02.htm>.

McGreal, R., Anderson, T., Babin, G., Downes, S., Friesen, N., Harrigan, K., et al. (2004). EduSource: Canada's learning object repository network. *International Journal of Instructional Technology and Distance Learning, 1*(3). <http://www.itdl.org/Journal/Mar_04/article01.htm>.

Nesbit, J., & Belfer, K. (2004). Collaborative evaluation of learning objects. In R. McGreal (Ed.), *Online education using learning objects* (pp. 138–153). New York: RoutledgeFalmer.

Nielson, J. (2003). *Ten usability heuristics*. <http://www.useit.com/papers/heuristic/heuristic_list.html>.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nurmi, S., & Jaakkola, T. (2005). Problems underlying the learning object approach. *International Journal of Instructional Technology and Distance Learning, 2*(11). <http://www.itdl.org/Journal/Nov_05/article07.htm>.

Nurmi, S., & Jaakkola, T. (2006a). Effectiveness of learning objects in various instructional settings. *Learning, Media, and Technology, 31*(3), 233–247. doi:10.1080/17439880600893283.

Nurmi, S., & Jaakkola, T. (2006b). Promises and pitfall of learning objects. *Learning, Media, and Technology, 31*(3), 269–285. doi:10.1080/17439880600893325.

Ohl, T. M. (2001). An interaction-centric learning model. *Journal of Educational Multimedia and Hypermedia, 10*(4), 311–332. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/8438>.

Oliver, R., & McLoughlin, C. (1999). Curriculum and learning-resources issues arising from the use of web-based course support systems. *International Journal of Educational Telecommunications, 5*(4), 419–435. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/8840>.

Parrish, P. E. (2004). The trouble with learning objects. *Educational Technology Research and Development, 52*(1), 49–67. doi:10.1007/BF02504772.

Reimer, K., & Moyer, P. S. (2005). Third-graders learning about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching, 24*(1), 5–25. <http://www.editlib.org.uproxy.library.dc-uoit.ca/p/18889>.

Schell, G. P., & Burns, M. (2002). A repository of e-learning objects for higher education. *E-Service Journal, 1*(2), 53–64. <http://eservicejournal.org/sites/default/files/schell_abs.pdf>.

Schoner, V., Buzza, D., Harrigan, K., & Strampel, K. (2005). Learning objects in use: 'Lite' assessment for field studies. *Journal of Online Learning and Teaching, 1*(1), 1–18. <http://jolt.merlot.org/documents/vol1_no1_schoner_001.pdf>.

Siqueira, S. W. M., Melo, R. N., & Braz, M. H. L. B. (2004). Increasing the semantics of learning objects. *International Journal of Computer Processing of Oriental Languages, 17*(1), 27–39.

Sosteric, M., & Hesemeier, S. (2002). When is a learning object not an object: A first step towards a theory of learning objects. *International Review of Research in Open and Distance Learning, 3*(2), 1–16. <http://www.irrodl.org/index.php/irrodl/article/view/106/557>.

Stevens, J. P. (1992). *Applied multivariate statistics for the social science applications* (2nd ed.). Hillsdale, NJ: Erlbaum.

Vacik, H., Wolfslehner, B., Spork, J., & Kortschak, E. (2006). The use of COCOON in teaching silviculture. *Computers and Education, 47*(3), 245–259. doi:10.1016/j.compedu.2004.10.009.

Van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Education Theory, Research and Development, 53*(3), 1042–1629. doi:10.1007/BF02504793.

Van Zele, E., Vandaele, P., Botteldooren, D., & Lenaerts, J. (2003). Implementation and evaluation of a course concept based on reusable learning objects. *Journal of Educational Computing and Research, 28*(4), 355–372. doi:10.2190/PYTN-HTG8-UMBD-3T6C.

Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: Computer mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications, 25*(3), 1–8.

Wiley, D., Waters, S., Dawson, D., Lambert, B., Barclay, M., & Wade, D. (2004). Overcoming the limitations of learning objects. *Journal of Educational Multimedia and Hypermedia, 13*(4), 507–521.

Williams, D. D. (2000). Evaluation of learning objects and instruction using learning objects. In D. A. Wiley (Ed.), *The instructional use of learning objects*. <http://reusability.org/read/chapters/williams.doc>.

Further reading

Nesbit, J., Belfer, K., & Vargo, J. (2002). A convergent participation model for evaluation of learning objects. *Canadian Journal of Learning and Technology, 28*(3). <http://www.cjlt.ca/index.php/cjlt/article/view/110/103>.