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 used and evaluation has occurred in higher education (e.g., Cochrane, 2005; Haughey & Murhead, 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 as “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).
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; Gadadonis, Sedig, & Liang, 2004; Koochang & 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 & McLaughlin, 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; Nisbet & Belfer, 2004; Oliver & McLaughlin, 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 & Churchill, 2004; Kay & Knaack, 2005; Lin & Gregor, 2006; Macdonald et al., 2005; Reimer & Moyer, 2005; Van Zele, Vandelaele, Botteldooren, & Lenaerts, 2003), multimedia used (Brown & Voltz, 2005; Nisbet & Belfer, 2004; Oliver & McLaughlin, 1999), and willingness to use WBLTs again (Kay & 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) focusing 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-Morto & Lopez, 2007; Schoner et al., 2005; Vacić, 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 & Jaakola, 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

One missing data), 11–17 years of age (Anderson & Krathwohl, 2001) 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.

3.1.2. 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) multiple-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

<table>
<thead>
<tr>
<th>Construct</th>
<th>No. items</th>
<th>Possible range</th>
<th>Actual range observed</th>
<th>Internal reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>5</td>
<td>5–35</td>
<td>5–35</td>
<td>r = 0.93</td>
</tr>
<tr>
<td>Design</td>
<td>4</td>
<td>4–28</td>
<td>4–28</td>
<td>r = 0.87</td>
</tr>
<tr>
<td>Engagement</td>
<td>4</td>
<td>4–28</td>
<td>4–28</td>
<td>r = 0.92</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Scale item</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-learning 1 – learn</td>
<td>.759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-learning 2 – feedback</td>
<td>.772</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-learning 3 – graphics</td>
<td>.723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-learning 4 – new concept</td>
<td>.798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-learning 5 – overall</td>
<td>.740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-design 6 – help</td>
<td>.596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-design 7 – instructions</td>
<td>.823</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-design 8 – easy to use</td>
<td>.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-design 9 – organized</td>
<td>.744</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-engagement 10 – theme</td>
<td>.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-engagement 11 – engaging</td>
<td>.783</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-engagement 12 – fun</td>
<td>.821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-engagement 12 – use again</td>
<td>.798</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scale item | Eigen value | PCT of VAR | Cum PCT |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.16</td>
<td>62.8</td>
<td>62.8</td>
</tr>
<tr>
<td>2</td>
<td>1.10</td>
<td>8.5</td>
<td>71.3</td>
</tr>
<tr>
<td>3</td>
<td>0.87</td>
<td>6.7</td>
<td>78.0</td>
</tr>
</tbody>
</table>
(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).

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 3

<table>
<thead>
<tr>
<th></th>
<th>Learning (students)</th>
<th>Design (students)</th>
<th>Engagement (students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning (teachers)</td>
<td>0.56**</td>
<td>0.52**</td>
<td>0.50**</td>
</tr>
<tr>
<td>Design (teachers)</td>
<td>0.46**</td>
<td>0.40*</td>
<td>0.65***</td>
</tr>
<tr>
<td>Engagement (teachers)</td>
<td>0.32</td>
<td>0.36*</td>
<td>0.25</td>
</tr>
</tbody>
</table>

* p < .05 (two-tailed).
** p < .01 (two-tailed).
*** p < .005 (two-tailed).
were consistent with the theoretical framework proposed by Kay and Knaack (2006a). Three types of validity were considered in this paper: construct validity, convergent validity, and predictive validity. For the remainder of the paper, we define 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. One could safely assume that these WBLTs include 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 examined 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. One could safely assume that these WBLTs include 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.

5.3.3. 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 Blooms 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 focusing on application and analysis, but not remembering or understanding.

A possible explanation for this result is that all WBLTs were pre-selected 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

#### Table 4

<table>
<thead>
<tr>
<th></th>
<th>Learning (students)</th>
<th>Design (students)</th>
<th>Engagement (students)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlations</strong></td>
<td><strong>0.01</strong></td>
<td><strong>-0.08</strong></td>
<td><strong>-0.04</strong></td>
</tr>
<tr>
<td><strong>(n = 418)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td>0.11</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>(n = 253)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>0.16**</td>
<td>0.12</td>
<td>0.16**</td>
</tr>
<tr>
<td><strong>(n = 418)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>0.37***</td>
<td>0.30**</td>
<td>0.31**</td>
</tr>
<tr>
<td><strong>(n = 87)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Remembering
Understanding
Application
Analysis

*Correlations among WBLT Evaluation constructs and learning performance measures.*

\[ r < .05 \text{ (two-tailed).} \]

\[ r < .005 \text{ (two-tailed).} \]

\[ r < .001 \text{ (two-tailed).} \]
remembering and understanding, regardless of their perceptions of learning quality, design, and engagement. On the other hand, WBLTs may need to be better designed 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 confirm this speculation, although it is clear that expanding the range of learning performance measures 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-test 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 as 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 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.

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 focusing 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 matter more when a WBLT addresses higher level concepts.

Appendix A. WBLT Evaluation Scale

<table>
<thead>
<tr>
<th>Learning</th>
<th>Strongly disagree 1</th>
<th>Disagree 2</th>
<th>Neutral 3</th>
<th>Agree 4</th>
<th>Strongly agree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Working with the learning object helped me learn</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2. The feedback from the learning object helped me learn</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3. The graphics and animations from the learning object helped me learn</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4. The learning object helped teach me a new concept</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5. Overall, the learning object helped me learn</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>Strongly disagree 1</th>
<th>Disagree 2</th>
<th>Neutral 3</th>
<th>Agree 4</th>
<th>Strongly agree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. The help features in the learning object were useful</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7. The instructions in the learning object were easy to follow</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8. The learning object was easy to use</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9. The learning object was well organized</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Engagement</th>
<th>Strongly disagree 1</th>
<th>Disagree 2</th>
<th>Neutral 3</th>
<th>Agree 4</th>
<th>Strongly agree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. I liked the overall theme of the learning object</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11. I found the learning object engaging</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12. The learning object made learning fun</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13. I would like to use the learning object again</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>


References


