CIMLE *Current Issues in Middle Level Education* (2015) 20 (1), 18-28

Connecting Math Website Evaluation to an Authentic Learning Activity for Teaching Candidates

Robert G. Ziegenfuss
Michael Smith
Shippensburg University

This article will discuss two teacher training functions: One function is to give the teacher candidates practice in evaluating currently available mathematics websites used in grades K-8 for mathematics instruction. The second function is the evaluation of data by teaching candidates of 13 commonly used math sites by middle and elementary teachers. Research and data collection are combined with the authentic activity of evaluating mathematics websites using critical review and evaluation tools. The usage of technology to assist in the education process was been increasing with the advent of more complicated technologies and relevant software (e.g., mathematics based websites). Creating new ideas on how to potentially aid learning is important, but so is the evaluation of such sites and programs in order to determine that they (e.g., mathematics based websites) are pragmatic, user-friendly, and able to make a positive impact.

Math Websites and Closing Achievement Gaps. Mathematics teaching candidates need to be aware of the technology currently being used by school districts, as well as those available for use but not currently being used. Frequently, young teachers will be called upon to close the gaps among different subgroups of their students by the district or by the public. Often the mathematics related computed assisted programs are used to either bring students up in performance on standardized tests or to fill the gaps between different subgroups of students identified by ethnic, racial, or socioeconomic status (SES) criteria. For example, in a review of studies of technology used for mathematics instruction by Cheung and Slavin (2013), "17% of the eighth graders eligible for free lunch scored at proficient or better, while 45% of middle class students scored this well…" (p. 88). School districts and parents demand that gaps in achievement be reduced. See Table 1 for some comparisons of student achievement.

Math Websites and Differentiated Instruction. Computer-based assignments can be a key component for differentiated instruction for students who need individualization or more reinforcement of key components (Ormrod, 2014). Newly adopted computer programs, such as Everyday Mathematics, introduce many ways to solve problems and use a quick-paced conceptual-based learning, leave some students behind. The spiral curriculum which comes back to topics each year in an ever shorter time frame also can be troublesome for the students who have learning disabilities while working in an inclusive environment (Wolfolk, 2014). Students who process slower than others are candidates for such individualized computer site programs. George (2005) said, "In differentiated classrooms, students work at different paces, sometimes exercising varied learning options, and they are assessed using indicators that fit their interests and needs" (Woolfolk, 2014, p. 10). Students who excel in mathematics often have the reverse issue of those students who struggle, in that they become bored quickly after understanding on the first explanation or on their own (Ormrod, 2014). Again, the computer site individual instruction can be used to extend the accelerated learner or have the learner become enriched in a new topic.

Analyzing Websites. Silius and Tervakari (2002) modified the work of Nielsen (1993) to develop basic categories for analyzing websites. The three main broad categories (i.e., utility, the value

Table 1

Mathematics performance by eighth grade students at proficient or better

Eligible for free lunch	Middle class students	African Americans	Hispanics	American Indians	Asian-Americans	Whites
17%	45%	12%	17%	18%	54%	44%

Note. National Assessment of Educational Progress, 2011 (Cheung & Slavin, 2013) added, and the pedagogical usability) were addressed in this study, but broken into smaller categories. To establish criteria for evaluation of the math websites, other studies were examined. For example, Zhang, Duke, and Jimenez (2011) asked these questions:

1. "Who wrote this and what credentials do they have?
2. Why was it written?
3. When was it written?
4. Does it help meet my needs?
5. Organization of the site?
6. To-do list for the future." (p. 2).

Under the value added, Silius and Tervakari (2002) identified organization, skills development, and testing as critical components. In this study, the value added components are embedded in the following criteria: Concept development, Different learning styles, Assessments and standards (i.e., Demonstration of Concepts). The second major category identified was related to usability and defined further as "…easy to learn, efficient… and subjectively pleasing" (Silius & Tervakari, 2002, p. 2). In this study, a similar criterion evaluated was "Activities, visual, and graphics." The comments of the teaching candidates were also examined regarding this criterion. The third broad category identified by Silius and Tervakari (2002) was pedagogical usability and was examined in this study through the authentic practice of logging onto the website and conducting the evaluation of each of the different mathematics websites. The ratings, along with open ended comments, reflected the usability of each site.

Using teaching tips related to developing long-term thinking and memory skills, an authentic learning experience was developed for the teaching candidates using the following concepts. The first teaching concept was "Begin with what the students already know" (Ormrod, 2011, p. 94). This concept holds true for college students and was applied. Teaching candidates had either worked on mathematics websites, or observed them used by their cooperating teacher. Most of the students had neither explored the websites in depth nor had they gone to the higher thinking level of actually evaluating the sites.

The second concept was "Communicate the message that students can and should make sense of the things they study" (Ormrod, 2011, p. 95). In this concept, merely reading about websites available or being actively used by school districts does not help the teaching candidates make sense of how they might be used in their own classrooms. Thus, by evaluating the websites in seven different criteria, this forced the teaching candidates to experience the sites much as their students will in class. Piaget (1964) wrote "…to know an object, to know an event is not simply to look at it…" (Woolfolk, 2014, p. 71). Teaching candidates cannot just view the mathematics websites and expect to know what and how to use the sites. Rating the different websites allows students to make critical reviews using the highest levels of thinking. Choosing specific categories for evaluation forces the teacher candidates to examine criteria critical to successful mathematics instruction which might be otherwise overlooked. In addition, moving about within the sites offers the teacher candidates navigational experiences similar to their students.

Evaluation Criteria

In order to objectively evaluate a number of similar items, a set of rigid criteria that apply to each item are necessary. Teaching concept: "Provide an overall structure that people can use to organize a complex body of information" (Ormrod, 2011, p. 95). This study was designed with pre-determined criteria for evaluation. Using literature and mathematics methods books, criteria were established to gauge the value of each website (DeWalle, 2007; Van De Walle, Karp, & Bay-Williams, 2011a).

In addition to developing rigid criteria, the criteria must be relevant and applicable to current standards. For this, another teaching concept was used: Content coverage and links sufficient to purpose? Is the website current with the state and Common Core State or applicable state standards? Although the teaching candidates may not be able to check standards for every state due to time purposes, the teacher candidates can certainly evaluate the website in which their certification will be issued and the Common Core State Standards, which has application in forty-eight states.

In a similar vein, the criteria must also be current, another teaching concept: Is the information outdated and current? With the trends toward concept building, do the mathematics websites engage the students in activities which will either help them understand the mathematics concepts or build toward understanding the basic mathematics concepts in multiple mathematics subject areas?

The overall appearance and construction reflected the teaching concept: Design: "Is there a logical and consistent structuring of the subject matter?" (Tweedle et al., 1998, p. 268). In evaluating apps, Jonas-Dwyer, Clark, Celenza, and Siddiqui (2012) ask: "…Is the navigation obvious or hidden?" (p. 55). The relevance to this category for the math websites falls under the ability for the teacher to focus learning toward specific math concepts, math standards, or review for students. Can teachers search the websites for specific Common Core State Standards or specific math concepts and skills?

A final teaching concept to examine was: "Readability: Are images [or] sound used appropriately…? Does the screen seem uncluttered…?" (Tweedle et al., 1998, p. 269). The first question used in their study relates to the ability to represent children in a diverse way, such as ethnicity, race, religion, geographical orientation, and urban v. suburban v. rural. The second question may be of concern for children with learning disabilities who need to have a screen which is uncluttered, easy to read and focused on one item at a time. The design portion addressed in their study as "readability" was addressed under the category Accommodation for English language learners and children with disabilities (Tweedle et al., 1998, p. 269). The issue of sound could not be addressed in the teacher candidate study due to the use of computer labs to conduct the website evaluations.

How do you evaluate websites being used by math instructors and school systems? Figure 1 identifies the categories, which were used for evaluation by the teaching candidates.

Research Goals

The goals of the research were to connect different aspects of teacher training:

1. Enhance teaching candidates' experience and awareness of math websites currently being used in grades kindergarten through grade 8, using an authentic activity.
2. Develop critical thinking skills among teaching candidates related to mathematics website evaluation criteria by summarizing the pros and cons and rating math websites.
3. Determine the highest rated mathematics websites across multiple criteria, as well as disaggregated by specific criteria.

Figure. 1. Criteria used to rate mathematics websites.

4. Summarize the pros and cons of each website in data form and comments.
5. Develop a pilot study which can be used for training teacher candidates, as well as collect evaluation data for future math websites and similar technology, such as apps and iPads.

Data were disaggregated according to individual websites and the seven criteria listed: Student friendly site navigation; Demonstration of mathematics concepts; Different learning styles addressed; Accommodations for English language learners and children with disabilities; Activities, visual, graphics; Assessment included; State standards applied. This research was exploratory in nature; therefore, there were no formal hypotheses.

Method

Sample

Elementary and middle level teaching candidates from Shippensburg University of Pennsylvania were asked to evaluate 13 different websites currently used by school districts for mathematics instruction, reinforcement, and enrichment or to "catch students up in preparation for standardized testing." A total of 1976 evaluations by teaching candidates were analyzed (88.7 % female/11.3 % male). Nearly all of the evaluations included the sex of the evaluator (N=1950 valid/1976 total), with only 1.3% being invalid, or 26 (i.e., missing/incomplete) of the total 1976 being eliminated from the study, as indicated by Table 2 when evaluating by sex. When not examining sex of the evaluator, the number of valid (i.e., complete) evaluations was 1947.

Table 2

Frequency table indicating valid and missing evaluations by number and percentage of sample

Females		Males		Incomplete		Total
Valid	Percent	Valid	Percent	Missing	Percent	
1729	87.5	221	11.2	26	1.3	1976

Note. Table represents frequency of evaluations by sex, as well as valid and missing or incomplete data, with percentages.

Materials and Procedures

All teaching candidates were to evaluate each of the 13 different mathematics-related websites. Each of the websites was evaluated during the same class over a three-year period. Data was collected during the semester prior to subjects' student teaching experience, thus giving them a certain amount of familiarity with the websites possibly being used by their cooperating schools. Data were sorted by website and disaggregated according to the seven criteria: Student friendly site navigation; Demonstration of mathematics concepts; Different learning styles addressed; Accommodation and recommendations for students with learning/ ELL; Activities, visual, graphics; Assessment included; State standards applied.

Student friendly site navigation. Can the K-8 grade students navigate the website easily with little interference from the teacher? Once the introduction to the site is completed, the students need to be able to go on the site and conduct their own practice or enrichment, allowing the teacher to conduct instructional activities with other students. Students do not know the credentials of the experts writing the materials for their websites; therefore, it is up to the teaching candidates or the classroom teachers to determine not only is the material included in the website worthy, but does it meet the needs of the students, and will the students be able to access necessary practice and learning materials on their own after a basic introduction.

Demonstration of mathematics concepts. The development of mathematics concepts during instruction is more important now than in the past. Although there is a focus on meeting the Common Core State Standards and providing student success in the high stakes testing among mathematical leaders, a push for developing conceptual understanding of the basic concepts underlying mathematics knowledge recently has been in the forefront of mathematics instruction.

Different learning styles addressed. Clearly the website format lends itself to a particular kind of learner. However, it is worth looking at the possibility if different websites make an attempt to address the multiple ways students learn.

Accommodation and recommendations for students with learning/ELL. Can students who are learning English navigate the website? Among the things to check is whether the website addresses a wide diverse population. Does the website depict all areas of diversity built into its site, such as pictures depicting children of different ethnic, racial, and gender groups? Are the different geographic areas such as urban, suburban, and rural depicted in the site?

Student activities, visual, and graphics. Do the graphics and visuals grab the students' attention and hold their interest? This criterion is particularly pertinent because students in grades K-8 are exposed to numerous websites, games, and multiple technologies. To keep students engaged and learning mathematics, the activities, visuals, and graphics must be of a high caliber.

Assessment. Does the website have assessment tools built into the website such that the teacher can identify what each student has learned? Are the assessment tools built around the standards and concepts identified by the state, national, or professional association standards, such as the National Council of Teachers of Mathematics? Are the results of the assessment tools present on the site? Though this measure/question was used by Tweedle et al. (1998) in a different way, the generalized question has relevance to the current study. How are children evaluated? What kind of output is given to the teacher? Are there opportunities for retesting or multiple evaluations on the same concept or standard? Are there different types of evaluations offered to the students using the website? For example, might they be required to do assessments in vocabulary, problem solving, error analysis, and writing, in addition to the traditional computation types of assessment? What kind of output does the student receive when working through an evaluation? Are the rewards for successful learning of the concepts and standards appropriate and relevant to the age of the children? Lastly, are the rewards valued by the children using the website?

State standards applied. With so many states getting involved with the Common Core State Standards, this is a key component for evaluation. The data collected for this evaluation research was conducted in a state where state standards were still being used.

Cost. Although cost was one of the criteria, it was eliminated from the study after the data was collected as many teacher candidates could not determine the cost from the website and thus omitted the evaluation.

Information was collected on sex, minor concentration, and level of intended teaching, though only sex was analyzed. Minor concentrations and level of intended teaching were not analyzed because the certification programs for the state changed during the collection of the data impacting the expectations for grade levels and the concentrations among the teacher candidates.

Once the invalid entries were removed, each site had between 146 and 150 evaluations across the seven different criteria as represented by Table 3. All seven criteria were rated from 1-10 with a 10 being excellent and 1 being poor. Each of the seven criteria was then summed into an overall rating, maximum score of 70, and had a Cronbach's alpha .83 which suggests a strong reliability. Criteria for comparison evaluations were predetermined and given to each evaluator in an Xxcel format to fill out.

Specific directions as to how to begin or where to start were not given. The approach was much like that found in a study by Pierson (2001) in her description of how one high tech teacher learns new things. Teaching candidates explored the websites, allowing them to reflect on their techniques when working with their students (Pierson, 2001). In addition, by limiting instructions and allowing teaching candidates to explore the sites on their own, the goal was for them to understand the navigation ease of the site, as well as experience the navigation difficulties students might endure as well.

Table 3.

Table providing descriptives information (e.g., Mean) for each of the 13 Mathematics Websites

Math Website	N Valid	Mean	Median	Minimum	Maximum	25%	50%	75%
Brainpop	149	8.51	8.71	2.86	10	7.93	8.71	9.57
StudyIsland	151	7.05	7.43	0	9.71	6	7.43	8.43
NCTM.org	151	7.71	7.71	3	10	7	7.71	8.61
Eduplace	151	7.81	8	3.43	10	6.86	8	9
AAAKnow	152	7.5	7.57	2.14	10	6.57	7.57	8.57
FigureThis	152	6.01	6.5	0	10	4.43	6.5	7.86
FunBrain	150	5.61	6	0	10	3.25	6	8
NLVM.usu	146	7.47	7.57	3.29	10	6.57	7.57	8.57
APlusMath	152	6.95	7.14	2	10	5.71	7.14	8.29
PBSKids.org	147	7.65	8	0	10	6.71	8	8.71
PrimaryGames	151	7.16	7.29	0	10	6	7.29	8.43
CoolMath4Kids	147	7.37	7.57	3	10	6.29	7.57	8.57
Ixl.com/math	148	8.12	8.43	2.86	10	73.32	8.43	9

Note. Table presents Name, Number of Valid Evaluations, Mean Overall Score, Median Overall Score, Minimum Overall Score, Maximum Overall Score, 25% Quartile, 50% Quartile, and 75% Quartile for each of the 13 Mathematics Websites

Data Collection

Data was collected in all mathematics methods classes for elementary or middle level education majors. There were no specific concentrations among the different teaching candidates although the data was collected such that comparisons could be made among the different concentrations, such concentrations as reading, special education, early childhood and mathematics. Data was collected over multiple semesters, not including summer or other mini-semesters. Each class was given one full class session (75 minutes) and the opportunity to finish at home if they did not finish during class. Not every student evaluated each site; hence some sites have a smaller N (see Table 3). The data was analyzed using SPSS.

Results

Medians, lower and upper quartiles were compared using each site's average overall rating, as well as each individual criteria (see Table 3). The results were displayed in box and whisker plots, which easily show the median along with the center 50% of the ratings. The box and whisker plots match what is often displayed for standardized test scores sent to school districts for high stakes testing results and were chosen for this data analysis to give a visual picture of the data and how the ratings were done by the teaching candidates. The output shown in this fashion mimics the data output for high stakes testing done at the state level. Looking at the data across the 13 different websites gives a quick picture of where the strengths and weaknesses were among the websites, as well as the disparity in the ratings.

Teacher candidates understand that the box represents the center 50% of the data with the vertical line within the box indicating the median. The two whiskers or the extensions outwards from the box represent the upper 25% of the data and the lower 25% of the data after all data has been sorted least to greatest. The output graphs give a quick comparison across different criteria such as the case in the website criteria comparisons or with specific math/reading standards. Like the ratings among the websites, when teacher candidates are employed they will have experience reading box and whisker plots and will be able to use the skill in analyzing student output from high stakes testing at the state level.

The 13 websites were compared against each other to test for differences in preference based on Overall Rating, in order to determine if their ratings were significantly different. The ANOVA was significant, F (12, 1897) = 12.62, p <.001, suggesting that the ratings were different. When comparing the overall rating of each of the 13 websites, Brainpop emerged as the highest rated site, with an average overall rating of 8.53 and Ixl.com/math was the second highest, 8.15. The site Funbrain was rated as the lowest with a score of 5.66 and Figurethis was rated as the second lowest, 5.97. An additional ANOVA was preferred, testing for a sex difference among the evaluations. The results approached significance (p = .11), as men and women seemed fairly consistent in their evaluations, with only minor differences for several sites, see Figure 2. Though Funbrain was rated as the lowest, a number of subjects reported difficulty accessing the site, which likely biased their evaluation of Funbrain (see Supplemental Material for more information). Due to the access difficulties with Funbrain, Figurethis, the second lowest rated website, was used for comparisons and Funbrain was dropped. When compared together, Funbrain and Figurethis were not significantly different, t (300) = 1.24, p=.22, suggesting they were "equally" the lowest. The three highest rated sites were Brainpop, t (299) = 8.98, p<.001, Ixl.com t (298) = 8.98, p<.001, and Eduplace, t (301) = 7.62, p<.001; each was significantly higher than both Figurethis or Funbrain, though Funbrain had technical difficulties.

Brainpop was rated as significantly higher on overall rating than Ixl.com/math, which verifies that Brainpop was rated as the highest program. In terms of the assessments, Brainpop was significantly higher on student-friendly site navigation, accommodation and recommendations for students with learning/ELL, and activities, visual, and graphics. Brainpop was marginally significantly higher on different learning styles addressed than Ixl.com/math. Ixl.com/math was significantly higher on state standards applied than Brainpop.

Overall Ratings. The difference of the ratings by the teacher candidates when comparing Brainpop to Funbrain (i.e., product numbers 1 and 7) was easy to see, with Brainpop being highly

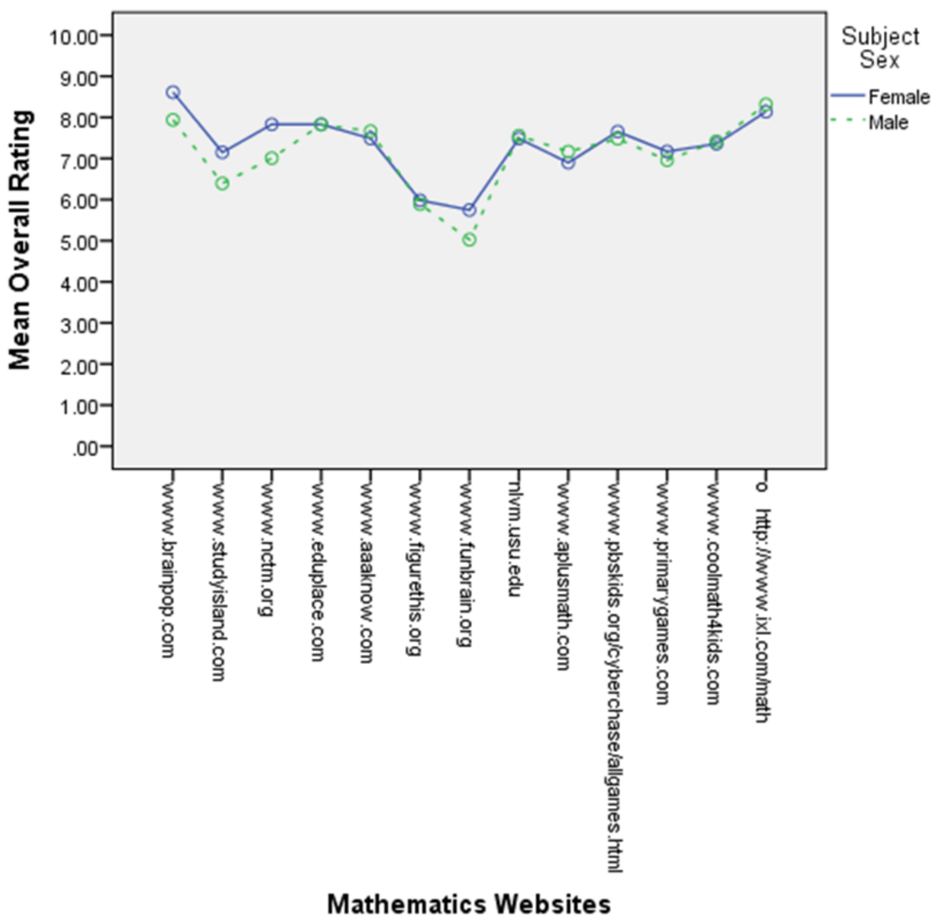

Figure 2. Graph of Mean Overall Score for Each of the 13 Mathematics Websites by Subject Sex.

favored. It also shows that the evaluators, in the case of Funbrain, were not nearly as consistent in their opinions with the center 50% of the data ranging from approximately 8 to approximately 3; whereas the center 50% of the evaluators for Brainpop were from 8 to a rating of 9.5. By visually scanning the box and whisker plots, it is clear that Funbrain had more disparity in its ratings than any other website. The large disparity in rating Figurethis could be due to the technical difficulty in accessing the website as expressed in the comment section of the evaluations. The second lowest total mean score rating was for Figurethis; this website also had the second highest amount of disparity in the center 50% of the evaluators (see Figure 3).

Student-Friendly Site Navigation. Two sites were overwhelmingly selected as the easiest to navigate according to the teacher candidate evaluators. Brainpop.com and PBSkids.org stood out above all other websites. In Brainpop.com and PBSkids.org, 75% percent of the teacher candidates rated these sites 9 or 10 out of 10 on student-friendly site navigation, suggesting a high degree of ease of navigation. Although two other sites, Coolmath4kids.com and Ixl.com/math, had equally high medians, the top 75% of the evaluators dropped to eight in both sites. Thus, all four sites were evaluated high for student-friendly navigation but Brainpop.com and PBSkids.org were consistently ranked the highest for this criterion (see Figure 4).

Figurethis.org and Studyisland.org showed the lowest median scores when comparing the rankings over the student friendly site navigation. This result was suspect because there was an issue with Studyisland.com. Some students had difficulty getting onto the site as it required signing up for a trial run or having a pass code to enter. This makes the low ranking of Studyisland.com for student friendly site navigation suspect, as many districts use it with a good bit of success and students seem to move about freely once there is a membership accepted by the school district.

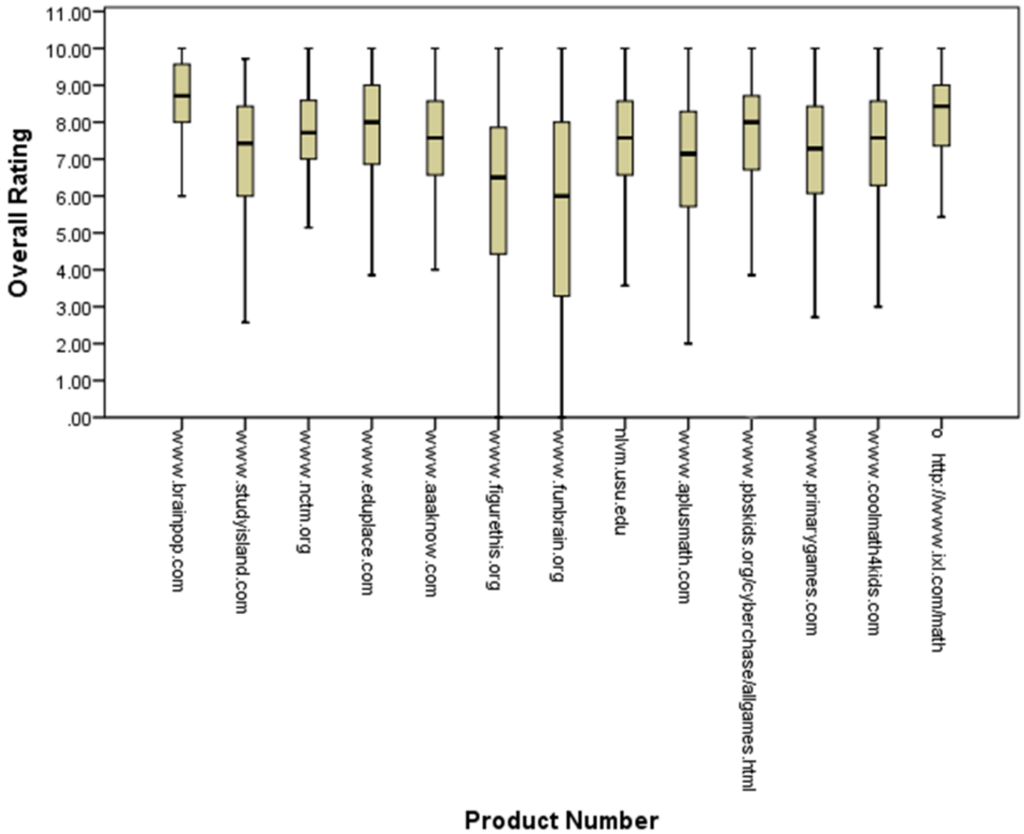

Figure 3. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by Overall Rating.

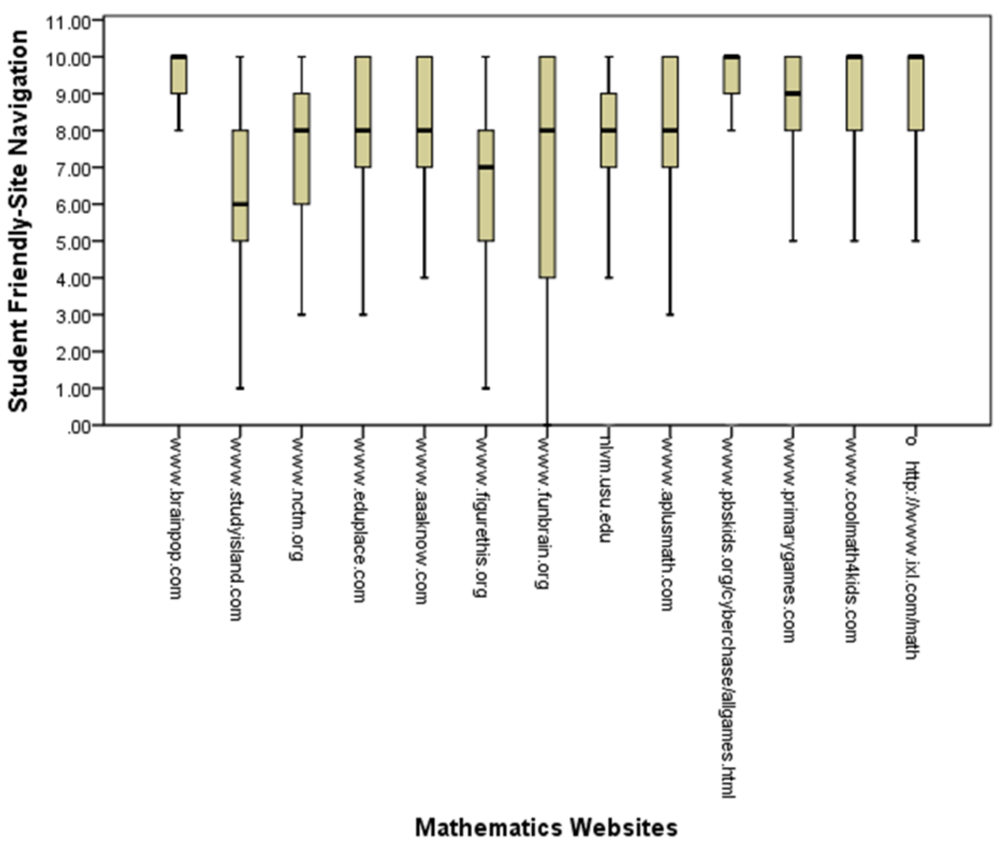

Figure 4. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by Student Friendly-Site Navigation.

Demonstration of Mathematics Concepts. There was not a single one or two websites which stood out in this area. A couple of factors might have played into this finding. Determining whether a website has met mathematics concepts is not always clear, particularly to students who are teacher candidates. It also takes much more in-depth evaluation to determine if these sites lead to concept development and it may be beyond the skills of teaching candidates. However, the candidates did determine two websites which they did not feel met this criterion (see Figure 5).

The two lowest rankings for demonstration of math concepts were Figurethis.com and Funbrain.org. The spread of the evaluations was the largest among all the site evaluations in this category, indicating that there was not consistency in the evaluation rating. Much as the overall ratings, the issue particularly with Funbrain.org could again be access.

Different Learning Styles Addressed. Perhaps the most significant finding of the evaluation of different learning styles addressed is that none of the websites addressed this criterion well. The nature of computer learning does not lend itself well to a variety of different learning styles but it is clear that none of the websites consistently addressed learning styles, according to our teaching candidates (see Figure 6).

Accommodations and recommendations for students with learning problems or those students who are English Language Learners. Eight of the 13 websites were not impressive, according to the evaluations, when judging the way websites made adjustments for students with learning disabilities or for those students learning the English language. Even among the five websites with some high evaluations, there

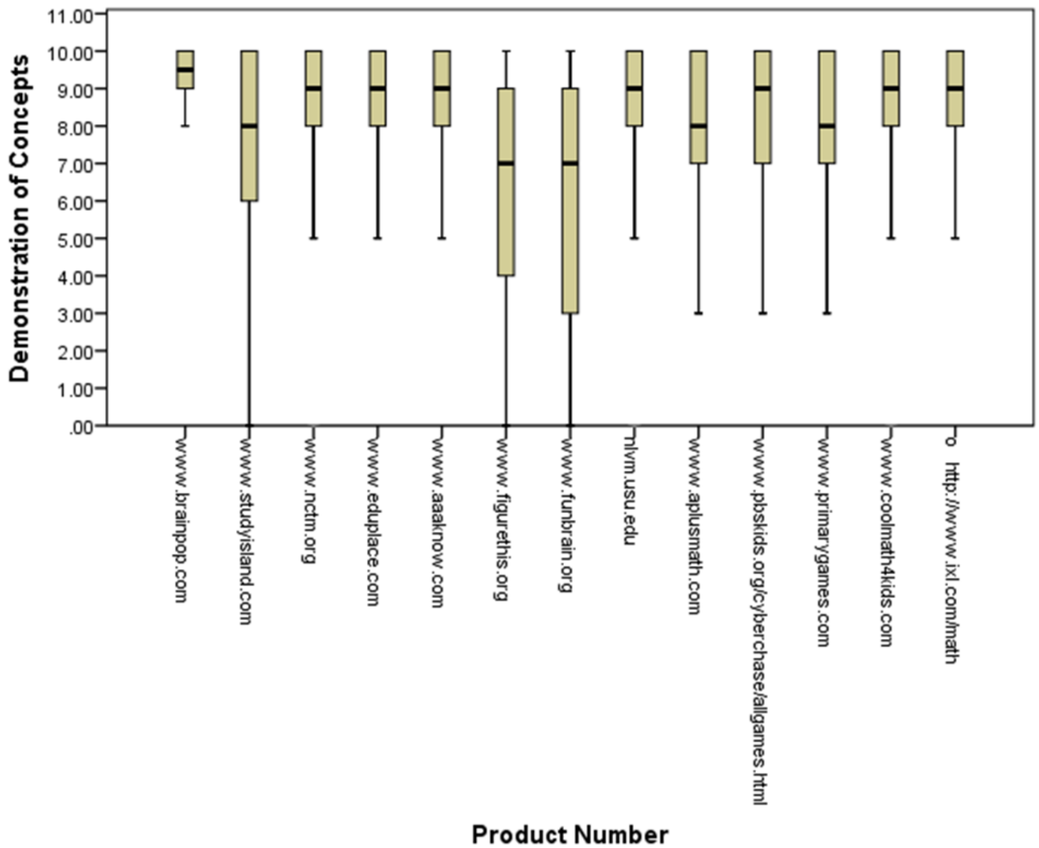

Figure 5. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by Demonstration of Mathematics Concepts.

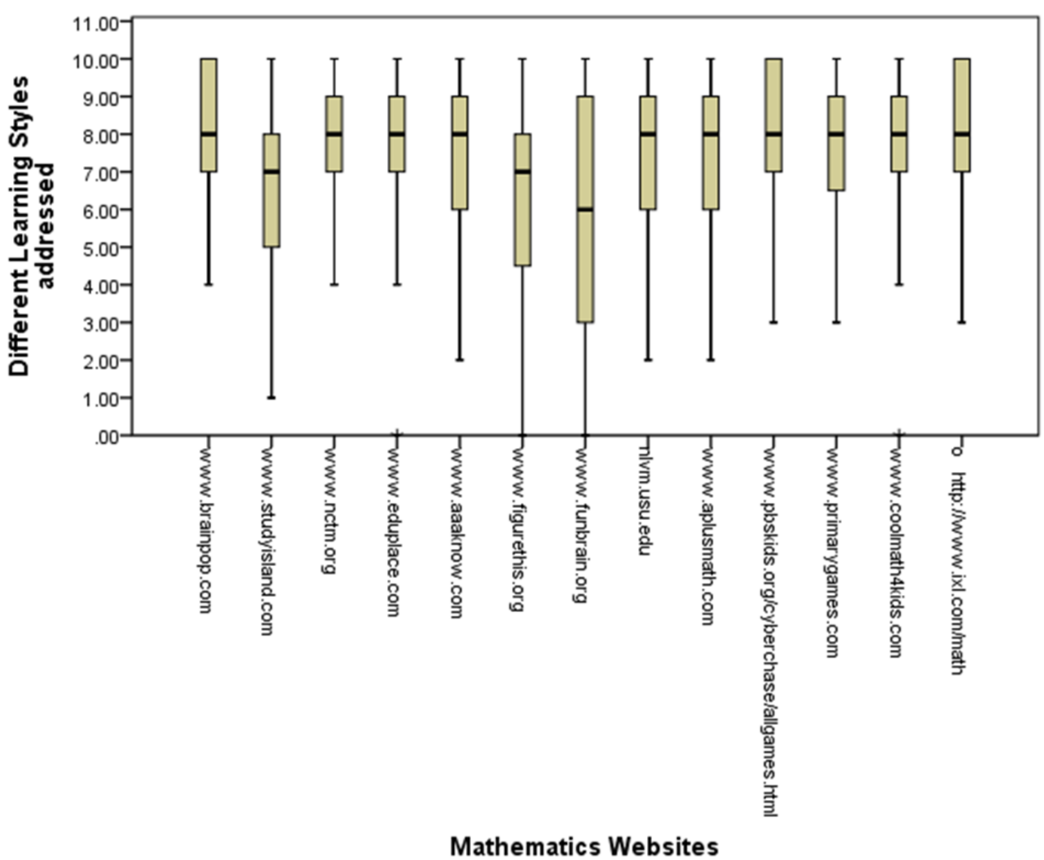

Figure 6. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by Different Learning Styles addressed by the sites.

was no consistency across all the teacher candidates. None of the websites had a median score above a ranking of 8 out of 10 (see Figure 7).

Activities, visuals and graphics. If math teachers are interested in providing excellent activities, visuals and graphics to their students, the teacher candidates rated Brainpop as the clear leader in this category with 75% of the candidates ranking this site as 10 out of 10. Where the median score was a 10/10, indicating that 50% of the teaching candidates rated the site 10, included Primarygames.com and PBSkids. Figurethis.org was the site rated the lowest (see Figure 8).

Assessment included. When considering the importance of evalating students, www.studyisland .com stood out above the other sites compared according to the teaching candidates. This combined with clear connections to the state level standards may be the reasons so many school districts have chosen to use this website although those reasons were not tested in this study. This is the only website with a median of 10 according to the evaluators (see Figure 9).

State standards. When considering state standards, four websites (i.e., Brainpop, Studyisland, Nctm.org, and Ixl.com) provided activities according to the standards. Seeing that Nctm.org provided the information on standards was not surprising, as this is the mathematics professional organization many states seek to provide assistance when setting up their state standards. Also, the organization was sought out to provide guidance for the Common Core State Standards for mathematics. The sites that did not rank as highly as the four mentioned may have information regarding state standards, but were not readily accessible during

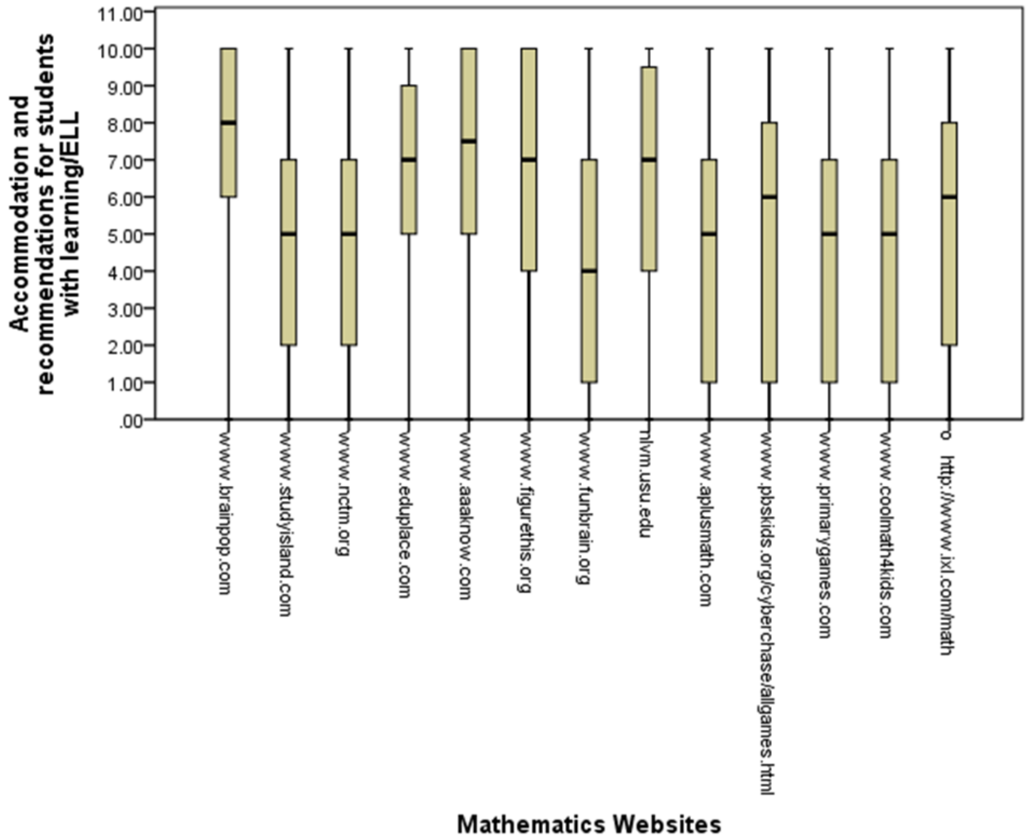

Figure 7. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by Accommodation and recommendations for students with learning disabilities or who are English Language Learners.

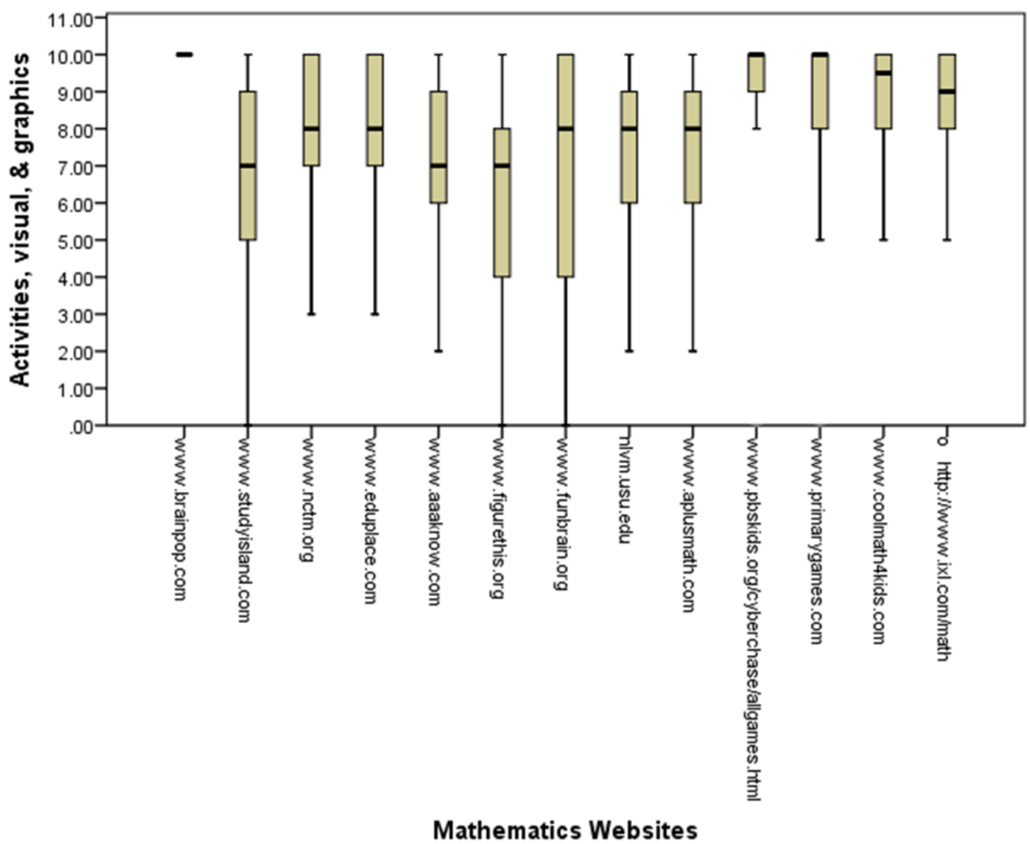

Figure 8. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by the quality of Activities, visuals, and graphics presented by the site.

the time frame committed by the students to each site evaluation (see Figure 10).

Qualitative Analyses – Results from student comments

Brainpop (N= 125). The real benefit of having students evaluate the websites showed up during the comment sections. The combination of holistic and analytic thinking was used in the study. Asking students first to rate individual characteristics of each site was an analytic activity geared to making specific analysis of pre-determined important criteria (Ormrod, 2014). The last column in their evaluation tool asked specifically for comments related to their overall opinion of the website or software. As many of the comments were different, there was a distinct pattern of students taking a close look at the different issues related to students using different websites-- the breadth covered, issues related to access of the site, and the availability for Spanish speaking and other English Language Learners. In the highest ranked website analyzed, Brainpop, specific words showed up among the different students making comments. Descriptor words by evaluators regarding Brainpop included: "easy to use; engaging; attractive; ESL link; Spanish offered; attractive; good videos; good animations; high interest." (See Appendix for additional positive review comments for Brainpop).

After reading the comments under Appendix, it is clear that certain things need to be emphasized from the instructor who is training teacher candidates. Do the children actually learn something or are they just interested in keeping the children entertained? Now, as the focus on national standards such as the Common Core Standards, "the fun activities" must enhance the learning and mathe-

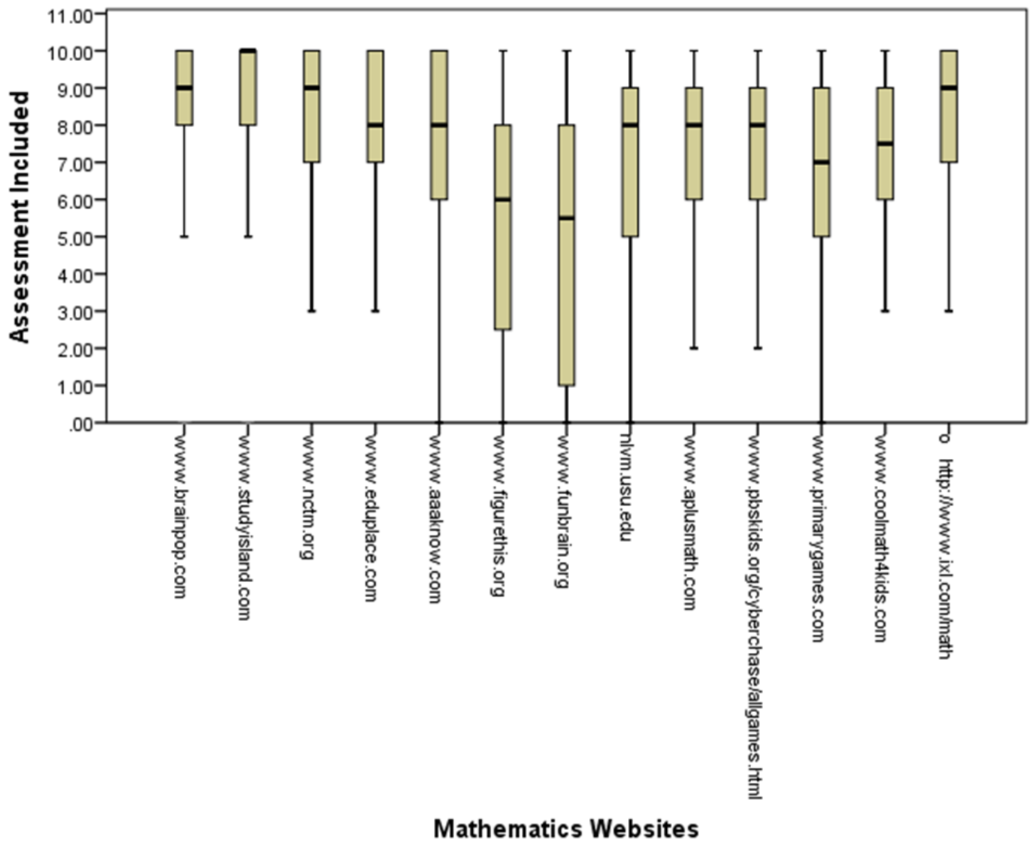

Figure 9. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by evaluative Assessment included in the site.

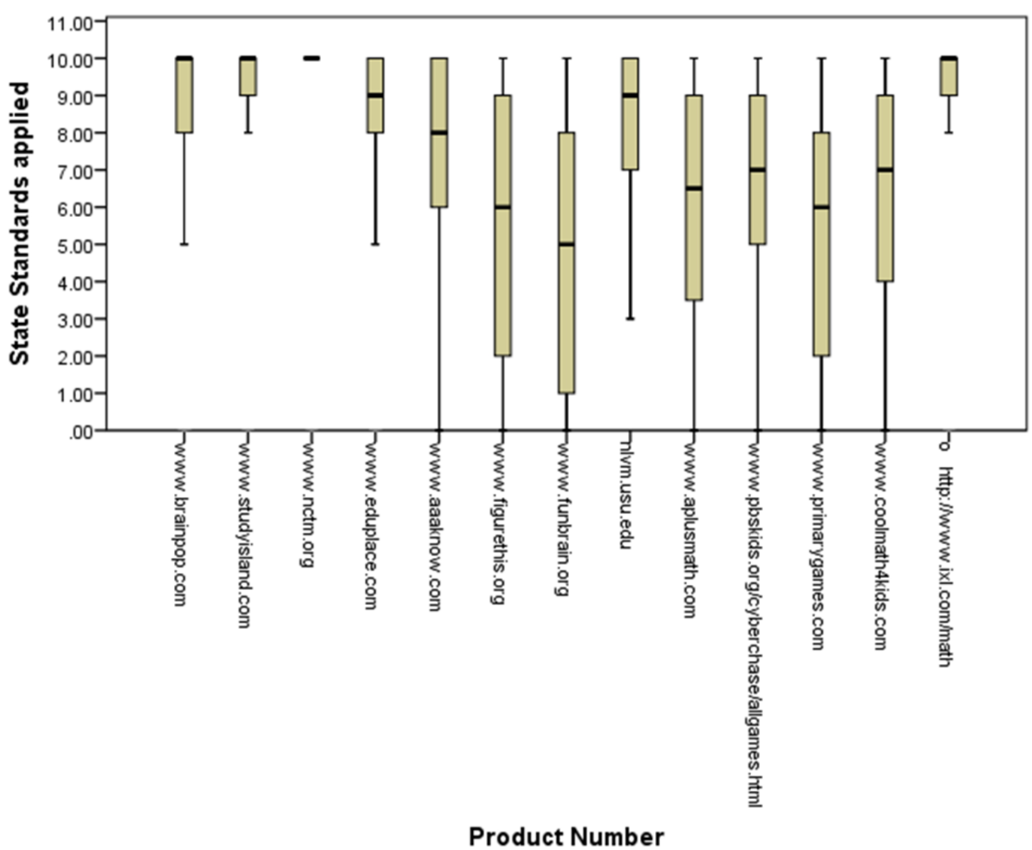

Figure 10. Box and Whisker plot of Teacher Candidates' evaluations of Mathematics Websites by how well the sites compared when States Standards or Common Core State Standards for mathematics were applied.

matical understanding. Professors training students hope to get the type of comment like the one made by this student, "I really find this website to be very useful. It has videos that are kid-friendly and that interest students and teaches them concepts as well. It has quizzes that can be great for assessments. The videos have words that you can read as well as sound that you can hear."

Issues. Although Brainpop was the most valued mathematics website evaluated, a common criticism came out. Having a membership or subscription was necessary to access much of the website which made it difficult for the teacher intern evaluators. Assessments, one of the critical components, were not accessible to students without the subscription. Other teacher candidates mentioned that should they ask students to access this at home, parents may not be willing or able to subscribe.

Figurethis (N=125). When the comments were examined for the web site Figurethis, most were found to be significantly negative which matches the statistical data. Many teaching interns complained of the number of links needed to be accessed. There were a few positive assessments of Figurethis as noted but positive evaluations were rare. The ones who liked the site were highly complementary (see Appendix). Although these were highly complementary, there were few like this throughout the comments. It was difficult to find positive comments like the ones found in the Appendix. It might be surmised that the students who found this site likeable were either not frustrated easily or had the knowledge and patience to connect to the many links offered.

The more likely comments were negative which were numerous and varied. They did have a common complaint which was that it was difficult because of the number of links. If teacher candidates get frustrated with accessing the area in which they wish to work in Figurethis, it is not much of a stretch to assume middle level students will be confused as well. Almost all of the teacher candidates expressed some variation of the following negative comments (see Appendix). There was an overall pattern in the comments which said that teacher candidates will not likely use this with their middle level students.

A strong benefit of the website evaluation activity was examining the different mathematics websites and looking at them in-depth using criteria advocated by researchers and practicing school-based educators. Teacher candidates were encouraged to make comments in the appropriate section. The intent was to encourage the students to determine whether they would use the website being evaluated in their own middle level classes.

Discussion. The reading of the comments allows one to see that there can be extensions to this evaluation activity after the data is analyzed. Much like asking students to interpret the data from high stakes testing, teaching candidates can be asked to target specific websites for a class that is diverse. For example, can certain websites be used for English Language Learners? Is there a particular website(s), which is better for reinforcing the Common Core State Standards in preparation for standardized testing?

General Discussion

Based on our sample and study, we have compared thirteen mathematics websites using seven different factors, as well as an overall rating, and found one (i.e., Brainpop) that was significantly higher rated than the other websites. However, with any research, there are potential limitations that may have affected the results. Reasons for disparity in website ratings may not be due entirely to the websites themselves but may be dependent on the individual evaluator's computer ability or his/her willingness to stick with the evaluation procedure. Whereas some students may have gotten frustrated and decided to rate a site with a low score, other evaluators may have continued to follow through until they learned the logistics of the site and then given the site a much higher evaluation. As one can see from the analysis of two low ranking websites, the disparate comments among the students indicated that such a scenario played out during the analysis. Fixing this problem in future evaluations may be difficult though necessary if a researcher or instructor is more interested in site evaluation than the authentic learning practice of traveling through a website. A time limit given to each site would take out the factor of one student spending more time than another, but might lead to a lack of sufficient time to truly investigate the website in-depth and would not account for differences in computer skills and website exploration skills. Therefore, an additional component to be studied may include the amount of time students are allotted to the computer for any given week. This is a structural component, as well as an attitudinal criteria, which depends on teacher willingness to accept new technology and how much it is valued by the teacher and the school.

A second factor impacting the evaluation data is that data was collected over several semesters and the websites are constantly being upgraded and changed. Thus, a website which may be evaluated poorly by the initial group may be rated highly by a later group of evaluators, or vice versa.

Future Studies

The work by Papadopoulos and Xenos (2008) leads to possibilities for future studies such as changing the methods, the populations, and in some cases the goals, while still continuing with the hidden goal of familiarizing teacher candidates with the technologies being used in mathematics middle level or elementary classrooms. The basic categories used in this study could be retained but altered slightly and asked of the student population using the websites instead of the teacher candidates. This may provide more detailed analysis of the success of the website from a students' point of view. As identified by Papadopoulos and Xenos (2008), the broad categories of content and technical completeness can be addressed but instead of the aesthetics, navigation, access, and usability, more student friendly questions can be designed appropriate for middle level students, which will better match the targeted audience. Likewise, with slight modifications, a similar study can target the latest mathematics apps being used with iPads. Important questions for future applied studies to examine include: What problems, if any, did you incur navigating the site? Were you able to find the area (s) that your teacher wanted you to work on in mathematics? How easy was it for you to move about the website? Did you learn anything from the using the material on the website? How did you know? Was the website easy to use, somewhat easy, difficult or very difficult to use? Did you like the graphics, pictures, games etc.? What did you like most about the website? What did you NOT like about the website?

In conclusion, this article discussed important findings for both evaluating mathematics applications, as well as pertinent criteria to focus on when developing new applications.

References

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9,* 88-113.

Hazma, M. K. (2002-2003). Web evaluation tool (WET): A creative web tool for online educators. *Journal of Educational Technology Systems, 31*(3), 307-329.

Jonas-Dwyer, D. R. D., Clark, C., Celenza, A., & Siddiqui, Z. S. (2012). Evaluating apps for learning and teaching. *iJET, 7*(1), 54-57.

Ormrod, J. E. (2011). *Our minds, our memories: Enhancing thinking and learning at all ages*. Boston, MA: Pearson.

Ormrod, J. E. (2014). *Educational psychology: Developing learners (*8th ed.). Boston, MA: Pearson.

Papadopoulos, T., & Xenos, M. (2008, August 29). *Quality evaluation of educational websites using heuristic and laboratory methods*. Proceedings 2nd Panhellenic Scientific Student Conference on Informatics, Related Technologies and Applications, pp. 42-54, August 29, Samos.

Pierson, M. E. (2001). Technology integration practice as a function of pedagogical expertise. *Journal of Research on Computing in Education, 33*(4), 413-431.

Silius, K., & Tervakari, A.M. (2002). *An evaluation of the usefulness of web-based learning environments: The evaluation tool into the portal of Finnish Virtual University*. DMI/Hypermedia Laboratory. Tampere University of Technology, Tampere, Finland.

Tweddle, S., Avis, P., Wright, J., & Waller, T. (1998). Toward criteria for evaluating websites. *British Journal of Educational Technology, 29*(3), 267-270.

Van De Walle, J. A., Karp, K. S., & Bay-Williams, J.M. (2013a). *Elementary and middle school mathematics: teaching developmentally* (8th ed). Boston, MA: Pearson.

Van De Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2013b). *Teaching student-centered mathematics, developmentally appropriate instruction for grades 3-5* (2nd ed). Boston, MA: Pearson.

Woofolk, A. (2014). *Educational psychology: Active learning edition* (12th ed). Boston, MA: Pearson

Zhang, S., Duke, N. K., & Jimenez, L. M. (2011) The WWWDOT Approach to Improving Students' Critical Evaluation of Websites. *Reading Teacher,65*(2), 150-158. doi: 10.1002/TRTR.01016