

Statistical Methods Used In Educational Technology Research 2012-2013

Nicola L. Ritter

ABSTRACT

This article provides a content analysis of the research methodologies used in quantitative and mixed-methods articles in the top five educational technology journals between 2012 and 2013. These articles represented a total of 32,131 sampling procedures and statistical techniques recorded from 1,171 articles – the largest research synthesis of research methodologies in field of educational technology to date. Results indicate quantitative methods continue to dominate the field as a whole, yet specific journals appear to favor certain research methods over others. Most authors did not report the type of sampling procedure used in their investigations (617 articles). Fewer researchers reported score reliability estimates using their own data – with only 420 articles reporting reliability coefficients. Findings also suggest few authors reported informationally-adequate statistics. Recommendations for best statistical practices and implications for the field of educational technology are discussed.

INTRODUCTION

Is educational technology research pseudo-scientific or filled with statistical blunders? Mitchell (1997) and Reeves (2000) suggested this over two decades ago. However, Mitchell and Reeves did not report any empirical evidence at the time to support these claims. The purpose of the current study is determine whether or not these claims are accurate. Research syntheses of research methodologies are published commonly in various fields, including education. Methodological reviews come in multiple forms, such as investigations of specific techniques published within a single journal or across multiple journals. For example, Willson (1980) reviewed articles published in *American Educational Research Journal* (AERJ) with respect to statistical methodology. More recently, disciplines within education have also begun to evaluate studies across multiple journals. For example, Warne, Lazo, Ramos, and Ritter (2012) reported the statistical techniques in five gifted education journals.

Educational technology researchers have also conducted methodological reviews. Most researchers identified the types of research methods used in educational technology literature (e.g., Koble & Bunker, 1997; Wang & Lockee, 2010). Other researchers extended their investigations beyond identifying research methods to synthesize the types of research designs (e.g., Cheung & Hew, 2009; Şimşek, Özdamar, Uysal, Kobak, Berk, Kiliçer, & Çiğdem, 2009). Likewise, reviews may investigate other methodological issues such as sampling method, reliability, and statistical techniques (e.g., Alper & Gülbahar, 2009).

LITERATURE REVIEW

Previous reviews adequately investigated the empirical nature, research methods, and research designs used in educational technology. Educational technology researchers captured information about the literature's empirical and non-empirical nature (e.g., Chen & Hirschheim, 2004; Farhoomand & Drury, 1999; Hrastinski & Keller, 2007). Empirical articles are published more often than non-empirical articles and the ratio between the two has remained constant over the past decade. Similarly, many researchers identified the research methods used in educational technology (Koble & Bunker, 1997; Rourke & Szabo, 2002; Wang & Lockee, 2010). Historically, quantitative methods dominated the field, yet today the field has seen a more balanced use of quantitative, qualitative, and mixed-methodologies. Lastly, researchers identified the experimental designs used, despite the lack of explicit reporting in the original article. Identifying trends in experimental designs are particularly difficult given the variation in classification schemes. Specifically, a synthesis among five reviews resulted in 14 different experimental design categories, with some overlapping others (Cheung & Hew, 2009; Koble & Bunker, 1997; Peterson-Karlan, 2011; Randolph, Julnes, Sutinen, & Lehman, 2008; Shih, Feng, & Tsai, 2008; Şimşek et al., 2009).

Despite the variety of previous reviews in educational technology, few researchers have reviewed sampling procedures, score reliability, or statistical techniques. Researchers found convenience samples to plague the field, yet the actual sampling procedures are often not directly reported (Alper & Gülbahar, 2009; Randolph et al., 2008). Likewise, the few researchers who reviewed score reliability information found poor reporting practices (Lee, Driscoll, & Nelson, 2004, 2007; Randolph, 2008). Finally, researchers who evaluated statistical techniques simply reported a brief list due to small sizes. For example, the list from Lee et al. (2004) was compiled from 47 articles, while the list from Lee et al. (2007) was compiled from 88 articles. The difficulty with the findings from the aforementioned studies is the small sample sizes and lack of discussion on these specific areas. As such more information and discussion is needed about sampling procedures, score reliability, or statistical techniques using a large sample size.

Researchers have a clear picture of the empirical nature, research methods, and research designs used in educational technology. Nevertheless, more information is needed about sampling procedures, score reliability, and statistical techniques used in the field. The need for this review arises due to numerous claims that the educational technology field is pseudo-scientific and lends itself to unsupported conclusions based on poor measure practice and statistical blunders (Mitchell, 1997; Reeves, 2000). However, little to no empirical evidence is used to support such claims.

Additionally, some researchers examined trends over time and across journals (e.g., Alper & Gülbahar, 2009; Chen & Hirschheim, 2004; Hrastinski & Keller, 2007; Koble & Bunker, 1997; Rourke & Szabo, 2002; Şimşek et al., 2009), while others did not observe trends (e.g., Cheung, & Hew, 2009; Farhoomand & Drury, 1999; Lee et al., 2004, 2007; Randolph, 2008, Randolph et al., 2008). Although these topics are explored to some extent, little is known about the statistical trends in educational technology. Moreover, few reviewed the educational technology field as a whole. For these reasons, a thorough synthesis of the field regarding the use of sampling procedures, score reliability, and statistical techniques is needed.

Methodological reviews benefit the publishing, research, and teaching communities. The current study can offer authors, editors, and reviewers insight to the publishing trends in high impact

educational technology journals. This study also identifies some of the strengths and weaknesses of current statistical techniques used in educational technology research. Lastly, educational technology doctoral programs can use this information to determine the types of statistical techniques their graduate students need to interpret and conduct research. By reviewing the use of sampling methods, score reliability and statistical techniques, in particular, the field may begin to determine the extent of the gaps other researchers claim that educational technology has “not yet held to the standards of experimental research in other fields of social science” (Lee et al., 2007, p. 40).

The purpose of the present article offers a comparable review of articles in educational technology to identify which sampling procedures and statistical techniques are used, and whether discernible trends emerged within journals and across the discipline over the past two years. Essentially, this review models Warne et al. (2012), who reviewed articles across multiple journals. Where other educational technology reviews are limited to specific areas, this study reviews the field as a whole. Moreover, this study moved beyond classifying research methods, and examined the sampling procedures and statistical techniques used in educational technology research.

METHODS

The present methodological review examined all articles published in the five influential educational technology journals over a two year period, from 2012 to 2013. The journals selected for review were based on the journals' impact factors using the *2011 Journal Citation Reports®* (JCR®) *Social Science Edition*. The journals in rank order were: *Computers & Education (C&E)*, *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, *British Journal of Educational Technology (BJET)*, *Australasian Journal of Educational Technology (AJET)*, and *Educational Technology Research and Development (ETR&D)*.

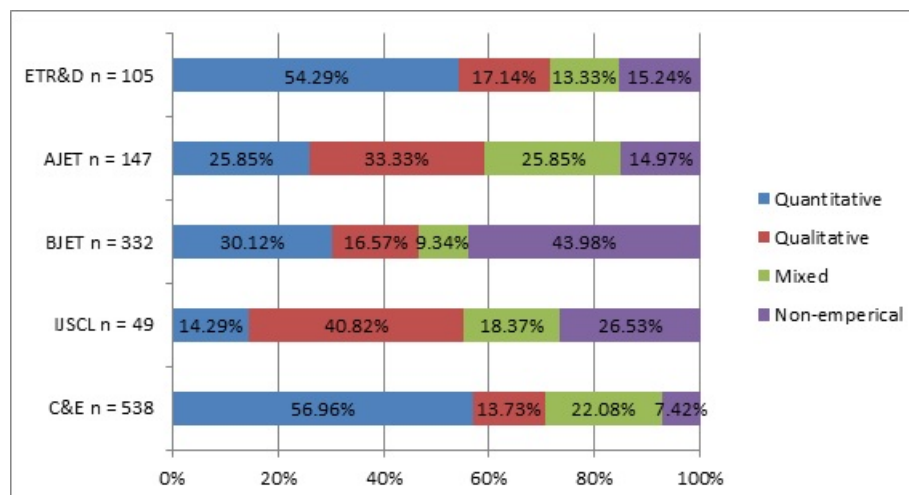
A coding scheme was created based on three sources: previous reviews (Skidmore & Thompson, 2010; Warne et al., 2012), publication standards (American Psychological Association, 2010; APA Publications & Communications Board Working Group on Journal Article Reporting Standards, 2008; Wilkinson & the Task Force on Statistical Inference, 1999), and meta-analytic coding suggestions (Cooper, 2010; Lipsey & Wilson, 2001). First, previous reviews were consulted, which led to an initial coding scheme. Based on the findings of previous reviews, it was observed that the original articles under review did not follow publication standards. As a result, the statistical reporting standards from the American Psychological Association (2010), APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008), and Wilkinson and the Task Force on Statistical Inference (1999) were appraised, which expanded the coding scheme to include specific statistics such as effect sizes. Lastly, works on meta-analytic coding were reviewed, which expanded the coding scheme to include the type of information to report in original articles to best synthesize information for future use (e.g. reporting standard deviations with means or accompanying means and standard deviations with correlation matrixes).

The coding sheet and definitions are located in supplementary materials section. Note, despite that some statistical analysis techniques are computationally equivalent to each other, the statistical technique was coded according to how the author of the article referred to the statistical technique. A total of 32,131 items from 1171 articles were recorded. The inter-rater reliability between two,

independent raters of a random sample of 35.44% of the articles ($n = 415$ articles) was Cohen's kappa = 0.99.

RESULTS

To compare the current study from previous methodological reviews, this study found empirical articles are favored over non-empirical articles, which mirror the findings of previous methodological reviews in educational technology. Among the 1,171 articles screened, 509 (43.47%) articles were quantitative, 216 (18.45%) were qualitative, and 211 (18.02%) were mixed-methods — a proportion similar to the findings from Hrastinski and Keller's (2007) review of educational technology articles from 2000 to 2004. The remaining 235 articles (20.07%) were non-empirical reports, such as editorials, book reviews, and other non-research articles. Figure 1 presents the type of research methods used across the five journals reviewed. Given the scope of this study, only quantitative and mixed-methods research reports were analyzed further. As such, while all 1,171 articles were screened, only 720 articles using quantitative or mixed research methods were further reviewed for sampling methods, score reliability, and statistical techniques.



Sampling procedures

Scholars previously suggested that educational technology research is plagued with convenience samples. In this study, most authors did not report the type of sampling procedure used. Specifically, authors of 617 articles (85.69%) did not report the type of sampling procedure used. Among the 720 articles, 22 articles (3.06%) indicated they used convenience sampling. Table 1 presents sampling procedures used across the five journals. Overall, the authors across all five journals rarely reported the type of sampling procedure used - less than 15% of articles ($n = 105$). Random assignment was rarely used in educational technology research, a finding similar to Lee et al. (2007). Authors of 146 articles (20.28%) randomly assigned participants to treatment groups.

Table 1: Sampling Procedure Reported in Education Technology Journals Organized by Journal

Sampling Procedure	Journal					Total
	<i>C&E</i>	<i>ijSCL</i>	<i>BJET</i>	<i>AJET</i>	<i>ETR&D</i>	
No sampling procedure reported	374 (87.79%)	15 (93.75%)	108 (82.44%)	62 (81.58%)	58 (81.69%)	617 (85.69%)
Convenience sampling	11 (2.58%)	0	4 (3.05%)	5 (6.58%)	2 (2.82%)	22 (3.06%)
Simple random sampling	14 (3.29%)	0	6 (4.58%)	7 (9.21%)	3 (4.23%)	30 (4.17%)
Database	12 (2.82%)	0	7 (5.34%)	1 (1.32%)	3 (4.23%)	23 (3.19%)
Purposive sampling	6 (1.41%)	1(6.25%)	2 (1.53%)	1 (1.32%)	1 (1.41%)	11 (1.53%)
Stratified random sampling	4 (0.94%)	0	1 (0.76%)	0	1 (1.41%)	6 (0.83%)
Cluster sampling	3 (0.70%)	0	1 (0.76%)	1 (1.32%)	1 (1.41%)	6 (0.83%)
Unspecified non-random sampling	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)
<i>Note.</i> The following sampling procedures were only used once: chain-referral sampling (<i>C&E</i>), stratified multi-stage cluster sampling (<i>C&E</i>), cluster-randomized sampling (<i>ETR&D</i>), probability proportional to size (PPS) in a multistage cluster sampling design (<i>ETR&D</i>).						

Score reliability

Scholars argue that researchers should report the reliability statistics from the dataset, because the interpretation of results is impacted by reliability. Thompson (1994) forewarned,

The failure to consider score reliability in substantive research may exact a toll on the interpretations within research studies. For example, we may conduct studies that could not possibly yield noteworthy effect sizes given that score reliability inherently attenuates effects sizes. Or we may not accurately interpret the effect sizes in our studies if we do not consider the reliability of the scores we are actually analyzing. (p. 840)

Despite these warnings, authors of only about half of the articles (n = 420) reported reliability data from their sample. This study mirrors Willson's (1980) findings 34 years ago. Moreover, authors of 16 articles (2.22%) invoked a reliability coefficient from previously reported datasets. Additionally, authors of 105 articles (14.58%) referred to reliability as an attribute of the test, instead of data at hand, by using phrases such as "reliability of the test" suggesting there may be misconceptions about reliability (Vacha-Haase, 1998). As Thompson (1992) noted,

This is not just an issue of sloppy speaking – the problem is that sometimes we unconsciously come to think what we say or what we hear, so that sloppy speaking does sometimes lead to a more pernicious outcome, sloppy thinking and sloppy practice. (p. 436)

This is not good practice because tests are not reliable. Gronlund and Linn (1990) explained,

Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself... Thus, it is more appropriate to speak of the reliability of the 'test scores' or the 'measurement' than of the 'test' or the 'instrument'. (p. 78).

Table 2: Reliability Methods Used in Educational Technology Journals, 2012-2013, Organized by Journal

Method	Journal					Total
	<i>C&E</i>	<i>IJSCL</i>	<i>BJET</i>	<i>AJET</i>	<i>ETR&D</i>	
Reliability induction	8 (1.88%)	0	3 (2.29%)	3 (3.95%)	2 (2.82%)	16 (2.22%)
Reported reliability for own data	256 (60.09%)	14 (87.50%)	74 (56.49%)	38 (50.00%)	38 (53.52%)	420 (58.33%)
"Reliability of the test" (not scores)	71 (16.67%)	0	12 (9.16%)	12 (15.79%)	10 (14.08%)	105 (14.58%)
Cronbach's alpha	201 (47.18%)	4 (25.00%)	54 (41.22%)	32 (42.11%)	22 (30.99%)	313 (43.47%)
Composite reliability	25 (5.87%)	0	6 (4.58%)	4 (5.26%)	1 (1.41%)	36 (5.00%)
Kuder-Richardson formula	13 (3.05%)	1 (6.25%)	2 (1.53%)	2 (2.63%)	4 (5.63%)	22 (3.06%)
Standardized Cronbach's alpha	2 (2.82%)	0	0	0	0	2 (0.28%)
Unspecified internal consistency	1 (0.23%)	0	3 (2.29%)	0	0	4 (0.56%)
Test-retest	6 (1.41%)	0	1 (0.76%)	0	1 (1.41%)	8 (1.11%)
Cohen's kappa	27 (6.34%)	6 (37.50%)	5 (3.82%)	4 (5.26%)	7 (9.86%)	49 (6.81%)
percent agreement	25 (5.87%)	4 (25.00%)	7 (5.34%)	1 (1.32%)	5 (7.04%)	42 (5.83%)
Krippendorff's alpha	4 (0.94%)	4 (25.00%)	0	0	1 (1.41%)	9 (1.25%)
Correlation between ratings	9 (2.11%)	1 (6.25%)	5 (3.82%)	1 (1.32%)	6 (8.45%)	22 (3.06%)
Free-Marginal Multirater Kappa	0	0	2 (1.53%)	0	0	2 (0.28%)
Intraclass correction coefficient	9 (2.11%)	1 (6.25%)	1 (0.76%)	1 (1.32%)	3 (4.23%)	14 (1.94%)
Unspecified inter-rater method	4 (0.94%)	0	0	1 (1.32%)	1 (1.41%)	6 (0.83%)
Alternate/ Parallel forms	6 (1.41%)	0	0	0	0	6 (0.83%)
Spearman-Brown	3 (0.70%)	1 (6.25%)	0	1 (1.32%)	0	5 (0.69%)
Guttman split-half	4 (0.94%)	0	0	0	0	4 (0.56%)
IRT-based reliability measure	9 (2.11%)	0	0	0	1 (1.41%)	10 (1.39%)

Note. IRT = item response theory; ICC = intraclass correlation coefficient; ROC = receiver operating characteristic curve; Little's MCAR test = Little's Missing Completely at Random test; Q-Q plot = Quantile-quantile plot. The following reliability statistics were used only once: Bland-Altman plot (C&E), Bayesian-based reliability measure (C&E), fidelity measures (C&E), H coefficient (C&E), McDonald's Omega (C&E), person separation reliability (C&E), Raykov's (2001) CFA-based method (C&E), squared multiple correlation (C&E), an unspecified reliability coefficient (C&E), rwg analysis (BJET), and Fleiss's kappa (ETR&D).

Table 3: Statistical Methods Used in Educational Technology Journals, 2012-2013, Organized by Journal

	Journal					
Method	C&E	IJSCL	BJET	AJET	ETR&D	Total
Descriptive statistics						
Means	343 (80.52%)	14 (87.50%)	86 (65.65%)	61 (80.26%)	62 (87.32%)	566 (78.61%)
SDs	298 (69.95%)	13 (81.25%)	76 (58.02%)	52 (68.42%)	59 (83.10%)	498 (69.17%)
Basic inferential statistics						
t tests	165 (38.73%)	5 (31.25%)	36 (27.48%)	32 (42.11%)	23 (32.39%)	261 (36.25%)
ANOVA	134 (31.46%)	10 (62.50%)	27 (20.61%)	16 (21.05%)	34 (47.89%)	221 (30.69%)
ANCOVA	40 (9.39%)	1 (6.25%)	10 (7.63%)	3 (3.95%)	7 (9.86%)	61 (8.47%)
MANOVA	19 (4.46%)	2 (12.50%)	7 (5.34%)	4 (5.26%)	7 (9.86%)	39 (5.42%)
MANCOVA	9 (2.11%)	1 (6.25%)	1 (0.76%)	0	0	11 (1.53%)
Correlation / regression						
Pearson's r	151 (35.45%)	6 (37.50%)	40 (30.53%)	15 (19.74%)	23 (32.39%)	235 (32.64%)
Spearman's rho	17 (3.99%)	1 (6.25%)	3 (2.29%)	3 (3.95%)	2 (2.82%)	26 (3.61%)
Point-biserial correlation	2 (0.47%)	0	0	0	1 (1.41%)	3 (0.42%)
Phi coefficient	2 (0.47%)	1 (6.25%)	0	0	0	3 (0.42%)
Kendall's tau	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Canonical correlation	0	0	0	1 (1.32%)	0	1 (0.14%)
Point-measure correlation	1 (0.23%)	0	0	0	0	1 (0.14%)
Multiple regression	48 (11.27%)	3 (18.75%)	12 (9.16%)	4 (5.26%)	7 (9.86%)	74 (10.28%)
Stepwise regression	12 (2.82%)	1 (6.25%)	5 (3.82%)	1 (1.32%)	0 (0.00%)	19 (2.64%)
Logistic regression	8 (1.88%)	0	2 (1.53%)	0 (0.00%)	1 (1.41%)	11 (1.53%)
Hierarchical linear regression	15 (3.52%)	0	4 (3.05%)	1 (1.32%)	3 (4.23%)	23 (3.19%)
HLM	9 (2.11%)	0	0	0	8 (11.27%)	17 (2.36%)
Data reduction						
EFA	23 (5.40%)	0	9 (6.87%)	4 (5.26%)	0	36 (5.00%)
PCA	31 (7.28%)	0	6 (4.58%)	6 (7.89%)	1 (1.41%)	44 (6.11%)
Unspecified	2 (0.47%)	0	1 (0.76%)	2 (1.32%)	0	5 (0.69%)
Structural Equation Modeling						
Path analysis	19 (4.46%)	0	6 (4.58%)	8 (10.53%)	0	33 (4.58%)
CFA / measurement model	52 (12.21%)	0	11 (8.40%)	7 (9.21%)	4 (5.63%)	74 (10.28%)
Structural model	32 (7.51%)	0	7 (5.34%)	2 (2.63%)	0	41 (5.69%)
Nonparametric statistics						
Pearson's X ²	46 (10.80%)	6 (37.50%)	12 (9.16%)	9 (11.84%)	7 (9.86%)	80 (11.11%)
Mann-Whitney U	28 (6.57%)	0	7 (5.34%)	1 (1.32%)	0	36 (5.00%)

Cluster analysis	15 (3.52%)	1 (6.25%)	4 (3.05%)	1 (1.32%)	3 (4.23%)	24 (3.33%)
Wilcoxon signed ranks	18 (4.23%)	1 (6.25%)	4 (3.05%)	0	0	23 (3.19%)
Kolmogorov-Smirnov test	10 (2.35%)	0	1 (0.76%)	1 (1.32%)	1 (1.41%)	13 (1.81%)
Kruskal-Wallis test	8 (1.88%)	0	2 (1.53%)	1 (1.32%)	0	11 (1.53%)
Fisher's exact test	4 (0.94%)	0	0	0	0	4 (0.56%)
Anderson-Darling test	2 (0.47%)	0	2 (1.53%)	0	0	4 (0.56%)
Friendman's X²	2 (0.47%)	0	0	0	0	2 (0.28%)
Kendall's W	0	0	0	2 (2.63%)	0	2 (0.28%)
McNemar's test	2 (0.47%)	0	0	0	0	2 (0.28%)
Sign test	1 (0.23%)	0	0	0	0	1 (0.14%)
Density Kernel estimation	1 (0.23%)	0	0	0	0	1 (0.14%)
Unspecified non-parametric test	1 (0.23%)	0	0	0	0	1 (0.14%)
Miscellaneous procedures						
Frequencies	259 (60.80%)	14 (87.50%)	79 (60.31%)	55 (72.37%)	42 (59.15%)	449 (62.36%)
Average variance extracted (AVE)	30 (7.04%)	0	7 (5.34%)	0	1 (1.41%)	38 (5.28%)
Levene's test	24 (5.63%)	0	2 (1.53%)	2 (2.63%)	5 (7.04%)	33 (4.58%)
Shapiro-Wilk test	7 (1.64%)	0	3 (2.29%)	1 (1.32%)	1 (1.41%)	12 (1.67%)
Power analysis	4 (0.94%)	0	1 (0.76%)	2 (2.63%)	5 (7.04%)	12 (1.67%)
Multicollinearity test	5 (1.17%)	0	2 (1.53%)	0	1 (1.41%)	8 (1.11%)
Social network analysis	4 (0.94%)	0	3 (2.29%)	0	1 (1.41%)	8 (1.11%)
Homogeneity-of-regression test	6 (1.41%)	0	1 (0.76%)	1 (1.32%)	0	8 (1.11%)
Mediation analyses	5 (1.17%)	2 (12.50%)	0	0	0	7 (0.97%)
Unspecified homogeneity of variance	5 (1.17%)	0	0	2 (2.63%)	0	7 (0.97%)
Bayesian statistics	4 (0.94%)	1 (6.25%)	0	0	1 (1.41%)	6 (0.83%)
Sequential analysis	3 (0.70%)	0	2 (1.53%)	0	0	5 (0.69%)
Wald test	3 (0.70%)	0	1 (0.76%)	0	1 (1.41%)	5 (0.69%)
Box's test	2 (0.47%)	0	1 (0.76%)	0	1 (1.41%)	4 (0.56%)
Item difficulty	3 (0.70%)	0	1 (0.76%)	0	0	4 (0.56%)
Item discrimination	2 (0.47%)	0	1 (0.76%)	0	0	3 (0.42%)
Square root of AVE	2 (0.47%)	0	0	1 (1.32%)	0	3 (0.42%)
Pillai's Trace	2 (0.47%)	0	0	0	1 (1.41%)	3 (0.42%)
Little's MCAR test	1 (0.23%)	0	0	1 (1.32%)	0	2 (0.28%)
Sobel test	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
ROC curves	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)
Fisher's Z transformation	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)

Q-Q plot	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Analytic Hierarchy Process	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Generalized Estimating Equation	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
F-test	2 (0.47%)	0	0	0	0	2 (0.28%)
Lawshe's content validity ratio	2 (0.47%)	0	0	0	0	2 (0.28%)
Welch's t-test	2 (0.47%)	0	0	0	0	2 (0.28%)
Mauchly's test	2 (0.47%)	0	0	0	0	2 (0.28%)
Mardia test	2 (0.47%)	0	0	0	0	2 (0.28%)
Bartlett's sphericity X² test	2 (0.47%)	0	0	0	0	2 (0.28%)
Unspecified linearity test	2 (0.47%)	0	0	0	0	2 (0.28%)
<i>Note.</i> ANOVA = analysis of variance; ANCOVA = analysis of covariance; MANOVA = multivariate analysis of variance; MANCOVA = multivariate analysis of covariance; HLM = hierarchical linear modeling; EFA = exploratory factor analysis; PCA = principal components analysis; CFA = confirmatory factor analysis. The following miscellaneous statistical methods were used only once: adjusted residuals tables (Z-score) (C&E), Box-Cox transformation (C&E), Carte & Russell's F-statistic (C&E), Cox-Stuart test (C&E), Curve estimation (C&E), Durbin-Watson-coefficient (C&E), Fuzzy Set Analysis (C&E), Harman's single factor test (C&E), Heterogeneity test: I² statistic (C&E), Holm-Bonferroni correction (C&E), Homogeneity test: Q statistic (C&E), Hosmer and Lemeshow Test of Reliability (C&E), Hotellings T² statistic (C&E), IRT-based estimate (C&E), Johnson-Neyman method (C&E), Kruskal-Wallis test with post hoc, Mann-Whitney U-test (C&E), Lagrange multiplier test (C&E), Likelihood ratio test (C&E), Mahalanobis distance (C&E), Mathieu's significance test (C&E), Multidimensional scaling (C&E), Multivariate normality (C&E), Ordinal logistic regression (C&E), Prediction interval (C&E), Prediction relevance (C&E), Propensity (C&E), Sensitivity analysis (C&E), Stochastic Frontier Regression (C&E), Unspecified normality test (C&E), Gini coefficient (IJSCL), log-likelihood ratios (G-tests) with post hoc H tests (BJET), Parallel analysis (BJET), post-hoc General Linear Hypothesis (BJET), Somer's d test (BJET), Tabachnick and Fidell's ratio validation (BJET), Unspecified heteroscedasticity test (BJET), Variance-stabilizing transformation (BJET), Jonckheere's test (AJET), Replication procedures (AJET), Bass modeling (ETR&D), Constant growth rate modeling (ETR&D), Factor score (ETR&D), F- test with Bonferroni correction (ETR&D), Ordinal Consistency (ETR&D), Proportional growth rate modeling (ETR&D), Stratified growth rate modeling (ETR&D), and Tversky's formula (ETR&D).						

The most often reported reliability coefficient was Cronbach's alpha, an internal consistency coefficient, appearing in 313 articles, almost half of the studies. The extensive use of Cronbach's alpha in educational technology research echoes general psychological literature (Hogan, Benjamin, & Brezinski, 2000), and gifted education (Warne et al., 2012). Authors of four articles (0.56%) computed a type of internal consistency coefficient but did not specify which type was calculated. This is troubling, because different estimates have different properties and may not be directly comparable across types. For a more detailed view refer to Table 2 that presents the type of reliability methods that were found across all five journals.

Statistical techniques

Table 3 presents the statistical techniques used across the journals reviewed. The following sections report on multiple methods used. Authors of articles often used more than one statistical technique.

Descriptive statistics

Most authors reported descriptive statistics. At times, means were reported without accompanying standard deviations or other variance-based statistics. The lack of reporting accompanying

standard deviations is acknowledged as poor practice (Thompson, 2006) and excludes articles from meta-analyses (Cooper, 2010).

Effect sizes

Recent publications stressed the importance of reporting effect sizes. Wilkinson et al. (1999) recommended that researchers should “*always* present effect sizes for primary outcomes” (p. 599, emphasis added). Moreover, Wilkinson et al., (1999), stated, “reporting and interpreting effect sizes in the context of previously reported effects is *essential* to good research” (p. 599, emphasis added). More recently, the APA manual (2010) embraces comparable language encouraging researchers to report effect sizes with the accompanying confidence intervals.

Despite scholars stressing the importance of reporting effect sizes, only 49.58% of the articles ($n = 357$) included at least one effect size. The current finding is lower than the findings from Randolph et al.’s, (2008) review, where they found that effect sizes were reported in 97.6% ($n = 120$) of computer science education research. Table 4 presents the types of effect sizes reported. The most often reported effect size was R^2 , which was found in 128 articles (17.78%) Other effect sizes were found and are listed in Table 4.

Occasionally, authors used Cohen’s (1992) benchmarks to interpret their effect size. However, this is not good practice. As Cohen (1988) himself noted,

These proposed conventions were set forth throughout with much diffidence, qualifications, and invitations not to employ them if possible... They were offered as conventions because they were needed in a research climate characterized by a neglect of attention to issues of [effect size] magnitude. (p. 532)

Among the articles reviewed, authors of 72 articles (10.00%) used Cohen’s benchmarks to interpret effect sizes. Of these 72 articles, 39 articles (54.17%) were found in C&E, three (4.17%) in *ijSCL*, nine in *BJET*, 14 (19.44%) in *AJET*, and 7 (9.72%) in *ETR&D*. Researchers also recommended that authors report confidence intervals (CI). Few authors reported CIs, which appeared in 69 articles (9.58%).

Table 4: Effect Sizes Reported in Education Technology Journals Organized by Journal

Effect Size	Journal					Total
	<i>C&E</i>	<i>IJSCL</i>	<i>BJET</i>	<i>AJET</i>	<i>ETR&D</i>	
R^2	89	1	20	9	9	128
Cohen's d	54	6	12	9	16	97
η^2	30	3	9	8	10	60
partial η^2	33	2	5	2	11	53
Variance explained	37	0	5	4	2	48
adjusted R^2	19	0	6	2	4	31
Unspecified effect size	16	0	2	3	3	24
r^2	10	0	1	1	2	14
Cramer's phi	8	2	1	1	1	13
Cohen's f^2	5	0	1	0	2	8
Pseudo- R^2	4	0	0	0	0	4
ω^2	2	0	0	0	0	2
Odds ratio	1	0	1	0	0	2
Hedges's g	1	0	0	0	1	2
Generalized Eta-squared (η^2_G)	1	0	0	0	0	1
Squared semi-partial (sR^2)	1	0	0	0	0	1
Canonical R^2 / R	0	0	2	0	0	2
κ^2	0	1	0	0	0	1
partial R^2	0	0	1	0	0	1

Basic inferential statistics

T-tests were the most common statistical method found after means and standard deviations. Moreover, ANOVA was the second most common inferential statistic, with 30.69% of articles (n = 221) using ANOVA. While MANOVA and MANCOVA was used less frequently.

Correlational and regression

The most commonly used correlational procedure is Pearson's r , appearing in 235 articles (32.64%). Other correlational techniques were rare. Multiple regression was the most common regression procedure (n = 74 articles; 10.28%). Other regression procedures were less common.

Advanced statistical techniques

Advanced statistical techniques are more accessible to researchers than ever before. However, few education technology researchers are taking advantage of more advanced quantitative analyses - comparable to the findings of Bulfin, Henderson, Johnson, and Selwyn (2014). For example, HLM allows researchers to analyze nested data such as data represented by three nesting levels (e.g., student, college, and university). HLM is rarely found in educational technology research, found in 17 articles (2.36%). Likewise, data reduction methods were rarely used.

Structural equation modeling (SEM) is a family of statistical techniques which extend beyond the constraints underlying the general linear model (Kline, 2011). Among the articles reviewed, 101 articles (14.03%) utilized at least one SEM method. Path analysis looks at relationships between

measured variables to other measured variables. Path analysis was seldom used with 33 articles (4.58%). While path analysis looks at relationships between measured variables, structural models looks at relationships between latent variables. Structural models were more common than path analysis, appearing in 41 articles (5.69%). Confirmatory factor analysis (CFA) and measurement models looks at relationships between measured variables to latent variables. CFA is the most common advanced statistical method used in educational technology with CFA/measurement models appearing in 74 articles (10.28%). Tests of invariance or result replicability that are often used in SEM and CFA were performed in 1 article.

Nonparametric statistics

Nonparametric statistics do not require assumptions of normality and associated probability distributions, which the aforementioned parametric methods do require. Nonparametric methods are more appropriate than parametric when assumptions are not met. Additionally, nonparametric statistics are often used when dependent variables are categorical in nature. The most common nonparametric statistical significance test was the Pearson's X^2 ; found in 80 articles (11.11%). There were many other techniques which did not fit well in previous sections. Note that some of these miscellaneous statistics are used concurrently with other statistical techniques. For example, frequencies were found in 449 articles (62.36%). Table 3 presents a comprehensive list of nonparametric and miscellaneous statistical techniques that were used.

DISCUSSION

This study found that quantitative methods continue to dominate the field as a whole, yet journals appear to favor certain research methods over others. As demonstrated in Figure 1, *C&E* and *ETR&D* favor quantitative methods more than other journals, while *ijSCL* favors qualitative methods. *AJET* has the most balanced representation of research methodologies, by slightly favoring qualitative methods. *BJET* has more non-empirical articles than any other journal – mostly due to numerous book reviews in each issue.

The choice of research method should be based on the researcher's questions. For example, the *ijSCL* focuses on in-depth collaboration within educational technology. Because of this, one would expect *ijSCL* to utilize more qualitative methods compared to general educational journals. Additionally, researchers recently discussed the harmony between quantitative and qualitative research, and encouraged mixed-method research (Johnson & Onwuegbuzie, 2004). As reported in Figure 1, this study suggests educational technology researchers are utilizing mixed-methods approaches.

Previous publications suggested that convenience samples plague the field, yet note that sampling procedures are often not directly reported (Alper & Gülbahar, 2009; Randolph et al., 2008). As seen in previous reviews of educational technology literature and Table 1, most authors did not report the type of sampling procedure used in their investigations. However, in most cases the descriptions of their participants suggested the sample was collected out of convenience. Nevertheless, convenience samples should not be considered poor research practice. Convenience samples are as helpful as other sampling procedures. However, when convenience samples are used, authors should adequately describe the sample for the reader to contextualize findings appropriately (see Gall, Gall, & Brog, 2007).

As reported in Table 2, this study found poor reporting practices just as Lee et al., (2004), Lee et al., (2007) and Randolph (2008) found. The lack of reporting reliability estimates continues to degrade educational research. Over 30 years ago, Willson (1980) observed,

Only 37% of the AERJ studies explicitly reported reliability coefficients for the data analyzed. Another 18% reported only indirectly through reference to earlier research... [and] unreported [reliability coefficients] in almost half the published research is... inexcusable at this late date. (p. 8-9)

In this study, approximately half of the authors reported reliability of the scores produced by their instrument, while fewer authors invoked a reliability coefficient from a previous study or test manual. Surely by now, the lack of reporting reliability coefficients for a given dataset is inexcusable. Accordingly, Wilkinson et al. (1999) stated, "if a questionnaire is used to collect data, [a researcher should] summarize the psychometric properties of its scores with specific regard to the way the instrument is used in a population" (p. 596). Authors should report evidence that the *scores* they are analyzing are reliable. The word *scores* is emphasized in the previous sentence because authors sometimes referred to reliability as an attribute of the test, instead of data at hand, by using phrases such as "reliability of the test." Verbiage relating reliability to a test instead of scores, suggests there may be misconceptions about reliability (Ritter, 2010). For a more thorough explanation about the importance of score reliability see Thompson (1992, 2003). It is important to note that all the articles reviewed may not need to report reliability coefficients based on the type of data collected. Nevertheless, the lack of reliability information is a weakness in educational technology research.

The results suggest few authors report informationally-adequate statistics. Certain statistics should be reported to effectively evaluate findings and to allow meta-analysts to conduct secondary research studies. Although this study identifies all statistical techniques used in educational technology between 2012 and 2013, two specific areas of concern are emphasized.

One area of concern is the tendency to report a mean without the SD about the mean. As reported in Table 3, among the authors of 566 articles reporting means, SDs were not reported for 68 of the articles (12.02%). In a similar vein, authors tended to report means of the instrument's items instead of variables used when performing SEM. The lack of reporting means and SDs of the variables used in the SEM hinders the ability of researchers in secondary studies to synthesize results from SEM analysis appropriately. Authors should always report means with their accompanying SDs and report means and SDs of the variables used when performing SEM.

LIMITATIONS

As with any study, there are some limitations to this study. First, while previous educational technology reviews are limited to specific areas, this study aimed to review the field as a whole by examining the top five educational journals between 2012 and 2013. Given there are at least 423 educational technology journals (cf., *2012 Cabell's Directories of Publishing Opportunities*) and only five, high-impact journals were selected, this study is limited in representing the entire field. By selecting the top five educational technology journals, the most impactful literature to the field was reviewed. Moreover, the top five journals ensure the level of quality research under review in a field with numerous publication outlets. Nevertheless, given that educational technology is a global issue,

future studies should examine other publications as well as publications in languages other than English.

Despite efforts to ensure the quality of the coding procedures, at times subjective judgment in classifying articles and interpreting authors' statistical explanations was required. Thus, if other researchers conducted a similar review, they would likely produce somewhat different results. In fact, Skidmore and Thompson (2010) found at times in other reviews when the same journal volumes of journals were analyzed by different authors different results were produced. Nevertheless, by having two coders and demonstrating sufficient inter-rater agreement ($\kappa = 0.99$), suggests that such problems were minimal in this study.

IMPLICATIONS

Although this study is descriptive, the results can provide authors, editors, and reviewers with insight to the publishing trends in high impact educational technology journals. The results also identify some of the strengths and weaknesses of current statistical techniques used in educational technology research. Additionally, decision-makers in educational technology doctoral programs can use this information to determine the types of statistical techniques their graduate students need to interpret and conduct research in the field.

There are three areas that educational technology researcher can improve on based on current reporting guidelines. Educational technology researchers should report the type of sampling procedure they used and provide adequate description of their samples so that readers can determine generalizability. Moreover, educational technology researchers should report evidence that the *scores* are reliable because of the impact reliability has on the interpretation of results. Lastly, researchers should report informationally-adequate statistics so that readers can evaluate findings appropriately. Is educational technology research pseudo-scientific or filled with statistical blunders? The field needs to decide.

References

- Alper, A. & Gülbahar, Y. (2009). Trends and issues in educational technologies: A review of recent research in TOJET. The Turkish Online Journal of Educational Technology, 8(2), 124-135.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. Educational Researcher, 35, 33-40.
- American Psychological Association (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: Author.
- APA Publications & Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology Why do we need them? what might they be? American Psychologist, 63, 839-851.
- Bulfin, S., Henderson, M., Johnson, N. F., Selwyn, N. (2014). Methodological capacity within the field of "education technology" research: an initial investigation. British Journal of Educational Technology, 45, 403-414.
- Chen, W. & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. Information Systems Journal, 14, 197-235.
- Cheung, W. S. & Hew, K. F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. Australasian Journal of Educational Technology, 25, 153-183.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.
- Farhoomand, A. & Drury, D.H. (1999). A historiographical examination of information systems. *Communications of the Association for Information Systems*, 1(5es), 1-27.
- Gall, M. D., Gall, J. P., & Brog, W. R. (2007). *Educational research an introduction* (8th ed.). Boston, MA: Pearson Education.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60, 523-531.
- Hrastinski, S. & Keller, C. (2007). An examination of research approaches that underlie research on educational technology: A review from 2000 to 2004. *Journal of Educational Computing Research*, 36, 175-190.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed-methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 14-26.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Koble, M. A & Bunker, E. L. (1997). Trends in research and practice: An examination of The American Journal of Distance Education 1987 to 1995. *American Journal of Distance Education*, 11, 19-38.
- Lee, Y., Driscoll, M. P., & Nelson, D. W. (2004). The past, present, and future of research in distance education: Results of a content analysis. *American Journal of Distance Education*, 18, 225-241.
- Lee, Y., Driscoll, M. P., & Nelson, D. W. (2007). Trends in research: A content analysis of major journals. In M. G. Moore (Ed.), *Handbook of Distance Education* (2nd ed., pp. 31-41). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Mitchell, P. D. (1997). The impact of educational technology: A radical reappraisal of research methods. *Association for Learning Technology Journal*, 5(1), 48-54. doi:10.1080/0968776970050108
- Peterson-Karlan, G. P. (2011). Technology to support writing by students with learning and academic disabilities: Recent research trends and findings. *Assistive Technology Outcomes and Benefits*, 7, 39-62.
- Randolph, J. J. (2008). A methodological review of the program evaluations in K-12 computer science education. *Informatics in Education*, 7, 237-258.
- Randolph, J., Julnes, G., Sutinen, E., & Lehman, S. (2008). A methodological review of computer science education research. *Journal of Information Technology Education*, 7, 135-162.
- Reeves, T.C. (2000, April). Enhancing the worth of instructional technology research through "design experiments" and other development research strategies. In Session 41.29, *International perspectives on instructional technology research for the 21st century*. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Ritter, N. L. (2010, February). Understanding a widely misunderstood statistic: Cronbach's alpha. Paper presented at the meeting of Southwestern Educational Research Association Conference 2010, New Orleans, LA (ED526237).
- Rourke, L., & Szabo, M. A. (2002). A content analysis of the Journal of Distance Education 1986-2001. *Journal of Distance Education*, 17(1), 63-74.
- Shih, M., Feng, J. & Tsai, C. C. (2008). Research and trends in the field of e-learning from 2001 to 2005: A content analysis of cognitive studies in selected journals. *Computers & Education*, 51, 955-967.
- Şimşek, A., Özdamar, N., Uysal, O., Kobak, K. Berk, C., Kiliçer, T., & Çiğdem, H. (2009). Current trends in educational technology research in Turkey in the new millennium. *Educational Sciences: Theory and Practice*, 9, 961-966.
- Skidmore, S. T. & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement*, 70, 777-795.

- Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development*, 70, 434-438.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837-847.
- Thompson, B. (Ed.) (2003). *Score reliability: Contemporary thinking on reliability issues*. Newbury Park, CA: Sage.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association
- Thompson, B. (2006). *Foundations of behavioral statistics*. New York, NY: Guilford Press.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58, 6-20.
- Wang, F. & Lockee, B. B. (2010). Virtual worlds in distance education a content analysis study. *Quarterly Review of Distance Education*, 11, 183-186.
- Warne, R.T. Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, 56, 134-149.
- Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. *Educational Researcher*, 9(5), 5-10.
- Wilkinson & the Task Force on Statistical Inference, (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.