Computers in Human Behavior 55 (2016) 159–166

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Research report

Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphClick, and DigitizeIt

Salih Rakap [a, *], Serife Rakap [b], Derya Evran [c], Oguzcan Cig [d]

[a] Department of Special Education, College of Education, Ondokuz Mayis University, Samsun, Turkey
[b] Dumlupinar Primary School, Carsamba, Samsun, Turkey
[c] Quantitative Social Science, Institute of Education, University College London, UK
[d] Department of Early Childhood Education, College of Education, Florida State University, Tallahassee, FL, 32306, USA

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form
10 September 2015
Accepted 12 September 2015
Available online 21 September 2015

Keywords:
Data extraction programs
Reliability and validity
Single-subject experimental research designs
UnGraph
GraphClick
DigitizeIt

ABSTRACT

The use of evidence-based practices in education has been gaining a lot of attention in recent years. Researchers often use meta-analyses to identify evidence-based practices. To conduct meta-analyses of studies employing single-subject experimental research (SSER) designs for the purpose of identifying evidence base for a practice, a necessary step is to obtain raw data from published graphs. One method for obtaining raw data from published SSER graphs is the use of computer programs specifically designed to extract data from graphs. Purpose of the present study was to examine the reliability and validity of three data extraction programs, Ungraph, GraphClick, and DigitizeIt, using 60 graphs obtained from 15 SSER studies focused on a practice. Three coders extracted data from the graphs using the three programs. Values extracted by each coder were compared to (a) each other (reliability) and (b) values reported in the original articles in which the graphs were obtained from (validity). Results showed that raw data from SSER graphs can be obtained reliably using all three data extraction programs and values obtained using the three programs are highly valid. These results suggest that researchers can use data extracted using these programs with a high level of confidences while conducting meta-analyses of studies employing SSER designs. Authors make recommendations for improving the accuracy of data extraction using the three programs.

1. Introduction

Movements towards the use of effective educational practices developed from scientifically based research in the classrooms to support development of all students including those with disabilities have necessitated the identification of evidence-based practices in education (Cook & Cook, 2013; Shavelson & Towne, 2002). Consequently, in recent years, several organizations and research groups have begun to identify such practices (Cook & Cook, 2013). For example, in 2002, the U.S. Department of Education awarded $18.5 million to the Institute for Education Science to establish the What Works Clearinghouse (WWC). The main goal of the WWC has been to identify reliable and scientifically based educational practices from which educators could make informed choices (Simpson, 2005).

The WWC and other organizations use meta-analysis, best-evidence synthesis, and systematic reviews of the literature to determine evidence-based practices (Rakap, Snyder, & Pasia, 2014). The process of identifying evidence-based practices involves evaluation of the quality of research design, quality and quantity of research studies, and magnitude of treatment effect (Cook & Cook, 2013). To determine magnitude of treatment effect, researchers often need raw data or other statistical estimates calculated using raw data (e.g., mean and standard deviation). Although group experimental and group quasi-experimental research studies usually provide the information necessary to calculate magnitude of treatment effect, the information to estimate the treatment effect is rarely provided in studies employing single-subject experimental research (SSER) designs, as the results of these studies are generally reported in the form of graphs (Kazdin, 2011; Kratochwill & Levin, 2015; Rakap, 2015; Wolery, Busick, Reichow, & Barton, 2010). Fig. 1 illustrates a graph obtained from a SSER study.

To conduct meta-analyses of studies utilizing SSER designs for the purpose of determining evidence base for a practice, a necessary step for researchers is to obtain raw data from published graphs. One method for obtaining raw data from published SSER graphs is the use of computer programs specifically designed to

* Corresponding author.
E-mail addresses: salih.rakap@omu.edu.tr, srakaptr@gmail.com (S. Rakap).

http://dx.doi.org/10.1016/j.chb.2015.09.008
0747-5632/

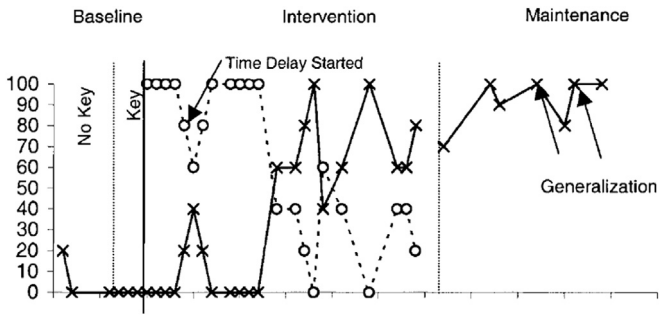

Fig. 1. An example of a graph obtained from a study employing a single-subject experimental research design. *Note*. The graph was obtained from Johnston, Nelson, Evans, and Palazolo (2003).

extract data from graphs. Purpose of the present study was to assess the reliability and validity of three such programs: UnGraph (Biosoft, 2004), GraphClick (Arizona-Software, 2008), and DigitizeIt (Bormann, 2012). There are several commercially available data extraction programs in the market. We chose to investigate these three programs among others because they have recently been used in meta-analyses of education research studies (e.g., Chenier et al., 2012; Dart, Collins, Klingbeil, & McKinley, 2014; Flower, McKenna, Bunuan, Muething, & Vega, 2014; Shadish et al., 2009; Sham & Smith, 2014; Stephenson & Carter, 2009).

1.1. Literature review

Two recent studies investigated reliability and validity of two such computer programs (Boyle, Samaha, Rodewald, & Hoffmann, 2013; Shadish et al., 2009). Shadish et al. (2009) used 91 SSER graphs that were randomly selected from 91 studies to assess reliability and validity of UnGraph. The reliability of the program was evaluated by comparing data extracted by two independent coders and the validity was evaluated by comparing data extracted by two coders with numerical descriptors (condition mean) of the graphs that the original authors may have presented in tables or text, using 44 of the 91 graphs. Authors computed Pearson product–moment correlations to assess the relationships between data extracted by two coders (reliability), and data extracted by two coders and values reported in original studies (validity). Results of the study showed high correlations between data extracted by two different coders (mean $r = .959$) and between means of values extracted by two coders and those reported in the original studies (range $r = .968 - .999$).

Boyle et al. (2013) replicated the Shadish et al. (2009) study with GraphClick. The authors used 191 graphs selected from the *Journal of Applied Behavior Analysis* to investigate the reliability of GraphClick and 15 graphs created with hypothetical data to evaluate the validity of the program. Three coders independently extracted data from these graphs. Pearson product–moment correlation analysis was used to assess (a) the relationship between data extracted by coders from the 191 graphs (reliability) and (b) the relationship between data extracted by coders from the 15 graphs with hypothetical data and the actual values used to create those graphs (validity). Results showed that GraphClick is a reliable and valid data extraction program with nearly perfect correlations for all analyses (reliability $r = .999$, $p < .0001$; validity $r = .999$, $p < .0001$).

1.2. Purpose of the research

Results of these two studies summarized above provide initial evidence about the reliability and validity of UnGraph and GraphClick for extracting data from graphs obtained from SSER studies. Two different research groups conducted these studies using different graphs and coders. In addition, graphs used in these studies to extract data were either randomly selected from published studies or selected from studies published in a journal between certain dates; they were not obtained from studies focused on a practice. We chose to use graphs obtained from a literature focused on a specific practice because data extraction programs are more likely to be used in the context of evidence-based reviews of identified intervention approaches (Kratochwill et al., 2010). Moreover, current literature does not include any study that investigates the reliability and validity of DigitizeIt (Bormann, 2012). Therefore, the purpose of this study was to assess the reliability and validity of Ungraph, GraphClick, and DigitizeIt programs by analyzing data extracted using these programs by the same coders using the same graphs obtained from SSER studies focused on a practice. In addition, we aimed to offer tips and guidance for researchers who wish to use one of these programs to extract data from SSER studies for the purpose of conducting meta-analyses of such studies.

2. Method

2.1. Sample

Graphs ($n = 60$) for the present study were obtained from 15 studies previously identified in a systematic review of literature focused naturalistic instructional approaches (see Snyder et al., 2015 for a description of search and article selection procedures). The original review included 40 SSER studies focused on a variety of naturalistic instructional approaches (e.g., milieu teaching, activity-based intervention, embedded instruction). For the present study, we only used graphs taken from studies focused on embedded instruction. These studies investigated the relationship between practitioners' or researchers' implementation of embedded instruction practices during ongoing activities of preschool classrooms and child learning outcomes. The 15 studies were published between 1984 and 2012 in 8 different journals. On average, each study contributed 4 graphs to the present study and each graph included 23.68 data points (range = 7–62).

2.2. Participants

Two doctoral students and a student with Master's degree (hereafter coders) extracted data from the graphs using the three data extraction programs. All three coders had experience in SSER and basic skills to use computers with Microsoft Windows or Mac OS X operating systems. One of the doctoral students had previous experience using all three programs. The other two coders did not have any experience using any of the three programs prior to this study. Experienced doctoral student was designated as the primary coder and therefore, extracted data from all 60 graphs using the three programs. The other two coders were designated as secondary coders and each extracted data from half of the entire set of graphs. Prior to coding for the present study, the doctoral student with experience using the three programs provided a brief training to the remaining coders. Training included a brief description of how to extract data using each program, model demonstrations of data extraction, and practice coding with feedback.

2.3. Materials

2.3.1. Data extraction programs

UnGraph, GraphClick, and DigitizeIt have fairly similar features. All three programs can import images in most file formats; handle linear, logarithmic or inverse scales; allow multiple data sets in the