Computers in Human Behavior 70 (2017) 518–522

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Is this the wave of the future? Examining the psychometric properties of child behavior ratings administered online

A.E. Pritchard [a, b, *], C.M. Stephan [a], T.A. Zabel [a, b], L.A. Jacobson [a, b]

[a] Department of Neuropsychology, Kennedy Krieger Institute, 1750 E. Fairmount Avenue, Baltimore, MD, 21231, United States
[b] Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD, 21287, United States

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form
19 December 2016
Accepted 12 January 2017
Available online 12 January 2017

Keywords:
Psychometric equivalence
Caregiver behavior ratings
Child psychology

ABSTRACT

Online data collection is increasingly utilized in the field of psychology. While there is a growing literature regarding the advantages of computer-based data collection, little is known regarding the psychometric comparability of measures across computer-based and traditional paper-and-pencil administration methods. This is especially true with regard to caregiver-report ratings of children's behavioral, academic, and emotional functioning. In this between-subjects study, the psychometric properties of caregiver behavioral, academic, and emotional functioning ratings for youth between the ages of 5 and 18 representing a mixed clinical sample were compared across computer-based (online group; n = 3831) and paper-and-pencil (P&P group; n = 514) administration formats. Groups differed on sex, age, and rater education distributions. Controlling for these variables, no statistically significant differences between groups were found in terms of symptomatology. In terms of psychometric characteristics, no group differences were found on inter-item reliability for any measure, and inter-item reliability for all scales across groups fell within the acceptable range. Similarly, comparisons of factor loadings across groups indicate marked consistency in the psychometric structure of all measures across administration formats. These results suggest that, regardless of administration method, the psychometric properties of caregiver ratings of childrens' behavioral, academic, and emotional functioning remain consistent.

1. Introduction

As technology becomes further embedded in everyday life, utilization of online data collection methods has become increasingly popular for the distribution of surveys and questionnaires. Online data collection offers many benefits in both the clinical and research realms. For clinical purposes the use of an online questionnaire offers advantages for patients, such as the ability to provide information to the clinician prior to the appointment, and at a time and location that is convenient for them (Buchanan, Johnson, & Goldberg, 2005). Relevant to both clinical and research operations, online questionnaires allow for the minimization of data entry and scoring errors as data acquisition, scoring, and analysis can be automated, which also allows for expedited feedback to the clinician or participant (Buchanan & Smith, 1999; Johnson, 2005). Similarly, the use of an online questionnaire has been found to reduce the number of missing data points, as online platforms can require respondents to answer specific questions before moving on (Stanton, 1998). Constraints placed on the structure of data can also be more effectively enforced using the online data collection format; such as content validation regarding response format (e.g., constraining reported dates to MM/DD/YYYY format). Costs are minimized with online administration of questionnaires, as is environmental impact (Naus, Philipp, & Samsi, 2009). In the research setting, online data collection enables studies to include a sample of increased geographic diversity and size (Johnson, 2005; Smith & Leigh, 1997) without concurrent increase in cost.

Research has also identified several drawbacks to the use of the online format for data collection. Chief among these are concerns about sample representativeness as a result of differences in access to and experience with computers and the internet. Higher levels of education are associated with greater access to the internet and less computer anxiety (Stanton, 1998); thus, use of an online format is

* Corresponding author. Department of Neuropsychology, Kennedy Krieger Institute, 1750 E. Fairmount Avenue, Baltimore, MD, 21231, United States.
E-mail address: Pritchard@kennedykrieger.org (A.E. Pritchard).

http://dx.doi.org/10.1016/j.chb.2017.01.030
0747-5632/

unlikely to yield a sample that is representative of the population as a whole with respect to sociodemographic factors. Familiarity with computers may also influence responding or performance, though research in this area has offered mixed findings (Luecht, Hadadi, Swanson, & Case, 1998; Mazzeo, Druesne, Raffeld, Checketts, and Muhlstein, 1991; Taylor, Jamieson, Eignor, & Kirsch, 1998) suggesting differential impacts of computer familiarity depending on task type (Russell, Ford, & Russell, 2015; de Beer & Visser, 1998). Equally important is the question of whether measures administered online are equivalent in terms of reliability and validity to measures administered via paper and pencil (P&P). Few measures have yet developed distinct norm sets for online and P&P administration formats; however, the psychometric properties of the two formats have been compared.

Examination of the comparability of the two formats has been focused largely on self-report and performance-based measures, offering mixed results. In the self-report domain, differences in anonymity of responses, the testing environment, and other extraneous variables that may influence responding have been hypothesized to contribute to differences in responding across formats; however, the influence of these factors seems to be minimal (Buchanan & Smith, 1999; Reips, 2000). Self-report measures of personality tend to demonstrate largely similar psychometric properties such as inter-item reliability and factor structure across administration formats (Buchanan, Goldberg, & Johnson, 1999; Johnson, 2000; Surís, Borman, Lind, & Kashner, 2007; Woolhouse & Myers, 1999; Rammstedt, Holzinger, & Rammsayer, 2004); however, even slight variations in factor structure across administration formats may have implications for measure scoring that must be addressed (Buchanan, 2000, 2002; Buchanan et al., 2005). Measures of clinical constructs such as anxiety and depression have also demonstrated similar but not identical psychometric properties across administration formats (e.g., Naus et al., 2009; Davis, 1999; Stones & Perry, 1997).

Thus, while administration format seems to have relatively little influence on responding in terms of the psychometric properties of a measure, the possibility of minor differences in factor structure and item loadings calls for evaluation of these characteristics on an individual, measure by measure basis in order to ensure that subscale, index, or standardized scores are comparable across formats. Comparability of caregiver ratings across administration formats has, as yet, been neglected in the research literature, and, as online survey dissemination becomes increasingly utilized, establishing psychometric comparability becomes even more critical. The goal of the present study is to evaluate the impact of administration format on the psychometric properties of three frequently used caregiver-report ratings of children's behavioral, academic, and emotional functioning. On the basis of the existing literature comparing self reports across administration formats, we hypothesize that caregiver ratings will be psychometrically equivalent across administration formats, though we expect to find slight differences in item factor loadings.

2. Method

2.1. Participants & procedures

Participants for the present study included caregivers of 4345 youth between the ages of 5 and 18 ($M = 10.16$ years, $SD = 3.24$, 64% male) who presented for evaluation in a large outpatient pediatric neuropsychology clinic between 2011 and 2015. As part of routine clinical care, caregivers completed a previsit questionnaire to provide behavior ratings for each child. This questionnaire was available to caregivers in both computer-based and paper-and-pencil formats. Both formats asked identical questions and offered the same response options; thus, the only difference was the mode of administration. After scheduling their clinic appointment, caregivers were mailed a packet of information which included a link to the computer-based questionnaire. If they had not completed the computer-based version of the questionnaire prior to their appointment, the determination of administration format was made based either on clinician or parent preference. Eighty-eight percent of the sample (n = 3831) completed the computer-based version of the questionnaire. Thus, the present study employs two naturally-occurring groups: computer-based respondents (online) and paper-and-pencil responders (P&P). Demographic characteristics of computer-based vs. paper-and-pencil respondents are detailed in the results section.

These data were studied under a waiver of consent, granted by the local institutional review board, which provided full approval for the study procedures. All procedures herein have been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Data were de-identified prior to analysis.

2.2. Measures

Measures were selected for inclusion in the previsit questionnaire on the basis of their value as screening instruments for the most prevalent emotional, behavioral, and academic conditions of childhood (e.g., Attention-Deficit/Hyperactivity Disorder, Oppositional Defiant Disorder, anxiety, depression, and learning disabilities). Measures that are free for use or available at low cost were prioritized when these measures demonstrated adequate psychometric properties and utility for screening.

2.2.1. ADHD Rating Scale – IV, Home Version (ADHD-RS-IV; DuPaul, Power, Anastopoulos, & Reid, 1998)

The ADHD-RS-IV is a parent-completed rating of symptoms of ADHD that closely corresponds to DSM-IV diagnostic criteria. The measure is comprised of Inattention (9 items) and Hyperactivity-Impulsivity (9 items) subscales. All questions are rated on a 4-point scale indicating how frequently symptoms are observed (0 = Never, 1 = Sometimes, 2 = Often, 3 = Very Often). Internal consistency fell above $\alpha = 0.86$ for all subscales. Adequate criterion-related, discriminant, and predictive validity have been established for the measure.

2.2.2. NICHQ Vanderbilt Assessment Scales, Parent Informant (Wolraich et al., 2003)

The Vanderbilt Assessment Scales are comprised of 55 items tapping a broad range of pediatric symptoms falling into the following diagnostic dimensions: ADHD, Oppositional Defiant Disorder (ODD), Conduct Disorder (CD), anxiety, and depression. For the purposes of the present study, a subset of items measuring ODD, CD, anxiety, and depression were utilized. ODD was assessed using 3 items: "loses temper," "actively defies or refuses to go along with adults' requests or rules," and "is angry or resentful." CD was assessed using 6 items: "is spiteful and wants to get even," "bullies, threatens, or intimidates others," "starts physical fights," "lies to get out of trouble or to avoid obligations (i.e., "cons" others)," "is physically cruel to people," "deliberately destroys others' property." Anxiety was assessed using 3 items: "is fearful, anxious, or worried," "is afraid to try new things for fear of making mistakes," and "is self-conscious or easily embarrassed." Depression was assessed using "blames self for problems, feels guilty," "feels lonely, unwanted, or unloved; complains that 'no one loves him or her'," "is sad, unhappy or depressed." All items are rated on a 4-point scale indicating how frequently symptoms occur (0 = Never, 1 = Sometimes, 2 = Often, 3 = Very Often). In full, the Vanderbilt