

Psychometrics, Ethics, Professional Standards, and Test Use: A Dilemma for British Psychologists Using the WISC-IV^{UK}

Gary L. Canivez
Eastern Illinois University
Charleston, IL, USA

Background

In preparing a manuscript reporting an investigation of the latent factor structure of the Wechsler Intelligence Scale for Children-Fourth UK Edition (WISC-IV^{UK}; Wechsler, 2004) with a large sample of Irish children referred for evaluation of educational difficulties (Watkins, Canivez, James, James, & Good, 2013), empirical studies examining the psychometric characteristics of the WISC-IV^{UK} with the standardization sample could not be found. In the 10 years since the WISC-IV^{UK} standardization data were collected there apparently are no published psychometric studies of the WISC-IV^{UK} standardization sample and I also learned that while the WISC-IV^{UK} was Anglicized and normed in the UK in 2003-2004, only a WISC-IV^{UK} *Administration and Scoring Manual* (Wechsler, 2004) was published. This manual contains a brief description of the modifications made to the WISC-IV for use in the UK, subtest administration directions and scoring guidelines, norms tables based on the representative standardization sample collected in the UK, and a brief comparison of raw score means and standard deviations from the UK and US samples. When a psychologist purchases the WISC-IV^{UK} they receive the test materials, the British version of the WISC-IV^{UK} *Administration and Scoring Manual*, and a copy of the US version of the WISC-IV *Technical and Interpretive Manual* (Wechsler, 2003). There is no technical and interpretive manual for the WISC-IV^{UK} to report evidence of reliability, validity, and utility based on the WISC-IV^{UK} standardization data. Confidence intervals provided in the WISC-IV^{UK} *Administration and Scoring Manual* are based on the standard errors of measurement derived from the reliability estimates obtained from the larger US WISC-IV standardization sample, not the reliability estimates from the WISC-IV^{UK} standardization sample. There are no estimates of short-term stability or interrater agreement in scoring the WISC-IV^{UK} and there are no validity studies of any kind reported for the WISC-IV^{UK}. Although the WISC-IV^{UK} *Administration and Scoring Manual* states “confidence in WISC-IV^{UK} score interpretation is based on the extensive US standardization study as well as the UK standardization, the close correspondence between these two sets of data, and the range of validation data reported in this Manual” (Wechsler, 2004, p. 284); reliability, validity, and utility of WISC-IV^{UK} scores remains unknown.

Because of the egregious omission of a technical manual reporting reliability and validity evidence for the WISC-IV^{UK} standardization sample; absence of technical reports of WISC-IV^{UK} reliability, validity, and utility; and the fact that there were no empirical psychometric studies of the WISC-IV^{UK} standardization sample published in the extant literature; I made a formal request to Pearson UK, the publisher of the WISC-IV^{UK}, to obtain the WISC-IV^{UK} standardization data (and WIAT-II^{UK} linking sample data) to conduct various validity studies for dissemination of results at professional conferences and peer reviewed journal articles. My initial request was rejected so I provided a more detailed request stating the importance of such research and that I had been previously provided WISC-IV standardisation data with WIAT-II linking sample from the US (for published results see Konold & Canivez, 2010) as well as WAIS-IV, WIAT-II, and WIAT-III data from Pearson. I also provided Pearson UK a copy of my most recent contract permitting my use of WAIS-IV, WIAT-II, and WIAT-III data collected by Pearson during WAIS-IV and WIAT-III standardization in the US (for published results see Canivez, 2013a). Again, I was denied access to the WISC-IV^{UK} standardization sample data. The specific reply I received was “Pearson's position on your request is that we do not

release data to external parties, so we will not be able to let you have the WISC–IV UK data. Our legal department will be in touch with you in due course.” Five months later I have yet to hear directly from Pearson’s legal department as to rationale why this request was denied. As a compromise, I requested the subtest correlation matrices and descriptive statistics by age group from the WISC–IV^{UK} standardization sample, which is statistical summary data that is customarily provided in all Wechsler intelligence scale technical manuals. I was again denied with the statement, “As per our initial decision, we will not be releasing our UK WISC–IV data in any form. I’m sorry for any inconvenience this causes, but our decision is final.”

Ethics and Codes

The British Psychological Society (BPS) ethical standards (*Code of Ethics and Conduct*, BPS, 2009) and practice codes (*Code of Good Practice for Psychological Testing*, BPS, 2010) directly and indirectly address issues of test use. Ethical standard 2.3 *Standard of recognizing limits of competence* states in part, “Psychologists should: (vi) Remain aware of and acknowledge the limits of their methods, as well as the limits of the conclusions that may be derived from such methods under different circumstances and for different purposes” (pp. 16-17). Further, the ethical principle of *Responsibility* includes the statement of values: “Psychologists value their responsibilities to clients, to the general public, and to the profession and science of Psychology, including the avoidance of harm and the prevention of misuse or abuse of their contributions to society” (p. 18). The ethical principle of *Integrity* includes the statement of values: “Psychologists value honesty, accuracy, clarity, and fairness in their interactions with all persons, and seek to promote integrity in all facets of their scientific and professional endeavors” (p. 21). More specifically, ethical standard 4.1 *Standard of honesty and accuracy* states in part, “Psychologists should: (iii) Be honest and accurate in conveying professional conclusions, opinions, and research findings, and in acknowledging the potential limitations” (p. 21). These ethical principles and statements can be applied to the situation where there is no technical information provided regarding the WISC-IV^{UK} standardization sample.

Further, the BPS *Code of Good Practice for Psychological Testing* (BPS, 2010) specifically addresses test use and the lack of technical information for tests. Under the *Procedures and Techniques* section, code 7 states “Use tests, in conjunction with other assessment methods, only when their use can be supported by the available technical information” (p. 2) and code 8 states “Administer, score and interpret tests in accordance with the instructions provided by the test distributor and to the standards defined by the Society” (p. 2). Finally, under the *Client Welfare* section, code 13 states “Give due consideration to factors such as gender, ethnicity, age, disability and special needs, educational background and level of ability in using and interpreting the results of tests” (p. 3). Without specific technical information regarding the WISC–IV^{UK} standardization sample, how can one abide by these codes? Which WISC–IV^{UK} scores, comparisons, or interpretive recommendations by the publisher have sufficient internal consistency, stability, or interrater agreement? Which WISC–IV^{UK} scores, comparisons, or interpretive recommendations by the publisher have sufficient evidence of validity? What evidence is there that the WISC-IV^{UK} structure allows for interpretation of various scores promoted by the publisher? There are many other questions that should be answered but without psychometric information from the WISC-IV^{UK} standardization sample or from independent UK samples it is rather difficult if not impossible for British psychologists to know. Although the US version of the WISC–IV *Technical and Interpretive Manual* (Wechsler, 2003) is provided to purchasers of the WISC–IV^{UK} can these results generalize sufficiently to the UK sample? Only analyses and publication of research with the WISC–IV^{UK} standardization sample and other independent UK samples will answer these and other critical questions.

Professional Standards

Professional associations such as the BPS, the International Test Commission (ITC), American Psychological Associations (APA), American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME), as well as others, are concerned with proper use of psychological tests and there are specific guidelines that address important responsibilities of both test users *and* test publishers (AERA, APA, NCME, 1999; ITC, 2000, 2010). The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and *ITC International Guidelines for Test Use* (ITC, 2000) both specify the importance for publishers and test developers to disclose critical psychometric features of tests so that those using such tests are fully informed of important aspects of test score reliability, validity, and utility; and to allow independent scrutiny of evidence. It is from this information, as well as from independent studies published in the extant literature, that clinicians are able to determine whether a specific test should be used in the situation they are considering, and if a test is used, which scores have sufficient reliability, validity, and utility to be interpreted and in what ways they should be interpreted. Without such information regarding the WISC-IV^{UK} standardization sample it would appear that British psychologists and others using the WISC-IV^{UK} in clinical practice are unable to make such judgments. Given these considerations, use of the WISC-IV^{UK} in clinical practice without consideration of key psychometric features of the WISC-IV^{UK} standardization sample would put British psychologists, and others using the WISC-IV^{UK}, at risk of violating ethical standards and codes of professional conduct. All tests and procedures must have key pieces of psychometric information disclosed so clinicians may make informed judgments regarding the purchase and use of specific tests with their clients and the motto of The Royal Society “Nullius in verba” (<http://royalsociety.org/about-us/history/>) seems particularly relevant.

Future Directions and Considerations

Several of my UK and Irish colleagues have been perplexed that Pearson UK has chosen to provide only the US version of the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003) in the sale and distribution of the WISC-IV^{UK} and have in the 10 years since collecting WISC-IV^{UK} standardization data failed to publish a separate WISC-IV^{UK} technical manual, technical report, or peer reviewed research article elucidating key psychometric information for the WISC-IV^{UK}. Even more disturbing is the blatant disregard for professional standards by denying independent analysis of standardization data, or even summary data (correlation matrices and descriptive statistics) commonly provided in technical manuals. Disclosure of important psychometric characteristics is of critical importance to clinicians using the WISC-IV^{UK}. Even if one were to accept psychometric information presented in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003) generalizing to the UK standardization sample there are numerous published studies that provide serious challenges to some of the claims and recommendations presented in the *WISC-IV Technical and Interpretive Manual* (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2013b; Watkins, 2006, 2010; Watkins et al., 2013; Watkins, Wilson, Kotz, Carbone, & Babula, 2006). In the meantime, analysis of WISC-IV^{UK} data gathered from UK clinical and research samples could provide some relevant psychometric information until such time Pearson UK discloses results from psychometric analyses on the WISC-IV^{UK} standardization sample or makes these data available for independent analyses. If there are psychologists who have collected WISC-IV^{UK} data from British clinical or research samples, publication of psychometric analyses using those samples are desperately needed to help address these glaring omissions. Given my interests in applied psychometrics research I would be happy to collaborate with anyone so interested. This research is obviously important and will help British psychologists follow Weiner’s (1989) sage advice that ethical psychologists must “(a) know what their tests can do and (b) act accordingly” (p. 829).

References

- American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (AERA, APA, NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15*, 417-424.
- The British Psychological Society. (2009). *Code of ethics and conduct*. Leichester, UK: Author.
- The British Psychological Society. (2010). *Code of good practice for psychological testing*. Leichester, UK: Author.
- Canivez, G. L. (2013a). Incremental validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment, 25*, 484-495. doi: 10.1037/a0032092
- Canivez, G. L. (2013b, July 29). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*. Advance online publication. doi: 10.1037/spq0000032
- International Test Commission (2000). *International Guidelines for Test Use*. [<http://www.intestcom.org>]
- International Test Commission (2010). *International Test Commission Guidelines for Translating and Adapting Tests*. [<http://www.intestcom.org>]
- Konold, T. R., & Canivez, G. L. (2010). Differential relationships among WISC-IV and WIAT-II scales: An evaluation of potentially moderating child demographics. *Educational and Psychological Measurement, 70*, 613-627. doi: 10.1177/0013164409355686
- Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment, 18*, 123-125. doi: 10.1037/1040-3590.18.1.123
- Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children-Fourth Edition among a national sample of referred students. *Psychological Assessment, 22*, 782-787. doi: 10.1037/a0020043
- Watkins, M. W., Canivez, G. L., James, T., Good, R., & James, K. (2013). Construct validity of the WISC-IV-UK with a large referred Irish sample. *International Journal of School and Educational Psychology, 1*, 102-111. doi: 10.1080/21683603.2013.794439
- Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children-Fourth Edition among referred students. *Educational and Psychological Measurement, 66*, 975-983. doi: 10.1177/0013164406288168
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2004). *Wechsler Intelligence Scale for Children—Fourth UK Edition*. London, England: Harcourt Assessment.
- Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment, 53*, 827-831.