The Educational and Developmental Psychologist

www.cambridge.org/edp

'It's a lot trickier than I expected': Assessment Issues and Dilemmas for Intern Psychologists

Linda Gilmore and Marilyn Campbell

Queensland University of Technology, Brisbane, Australia

Original Article

Cite this article: Gilmore L and Campbell M (2019). 'It's a lot trickier than I expected': Assessment Issues and Dilemmas for Intern Psychologists. *The Educational and Developmental Psychologist* **36**: 3–7. doi: 10.1017/edp.2019.3

Received: 1 December 2018
Revised: 8 March 2019
Accepted: 12 April 2019
First published online: 31 May 2019

Keywords:
psycho-educational assessment; issues; dilemmas; training; intern psychologists

Author for correspondence:
Linda Gilmore, Email: l.gilmore@qut.edu.au

Abstract

Beyond the basic information provided in foundation texts and test manuals, there is a notable scarcity of professional literature addressing complex issues and dilemmas in assessment. At times, psychologists need to make decisions based on their professional experience or personal preferences, and it is possible that some of these decisions are made without adequate justification or reflection. Intern psychologists may be confused by the differing advice they are given by university staff, supervisors and experienced practitioners. Our aim in this project was to identify specific issues in psychoeducational assessment that need clarity, guidance and reflection, not only during postgraduate training, but also across the professional lifespan. Focus groups with educational and developmental psychology interns revealed confusion about aspects of test administration and interpretation, as well as a number of challenging professional issues. The findings have implications for trainers and supervisors, as well as for experienced practitioners.

Intern psychologists who hold provisional registration and are undergoing supervised practice, either as part of a postgraduate degree in psychology or in the workplace in Australia, generally face a steep learning curve. There are many skills to be mastered in a short time. Following university undergraduate and fourth year coursework that contains very little, if any, client contact, interns are suddenly faced with real cases — usually a welcome experience after years of assignments and examinations, but also one that can be challenging and stressful.

Assessment is arguably one of the most demanding areas of practice for intern psychologists. It requires not only a sufficient background in psychometrics, but also a high level of familiarity with test administration, scoring and interpretation. Interns have rarely had any previous hands-on experience with tests, and they may need to quickly become competent in a number of complex instruments. Within masters' programs in Australian universities, basic training is provided, often in a robust and structured manner (e.g., Gilmore & Campbell, 2009). But for interns who go straight into the field for their supervised practice, the quality of training in test administration is variable, and may be patchy or inadequate.

Although non-psychologists (e.g., occupational therapists) are permitted to use some psychological tests (e.g., the Bayley Scales of Infant Development) provided they have appropriate training and supervision, complex psychological assessment is the exclusive domain of psychologists. In particular, intelligence testing is restricted to psychologists, and proficiency in the administration and interpretation of intelligence tests is considered to be a core competency in many specialisations within the profession, including educational and developmental psychology. Because test scores potentially have significant implications for decisions about educational placements and eligibility for various services, accurate administration, scoring and interpretation are imperative. Errors may have significant consequences, including misclassification of individuals and loss of public confidence in the profession (Belk, LoBello, Ray, & Zachar, 2002). Rigorous training in assessment is thus highly regarded and valued by the profession (Butcher, 2006; Krishnamurthy et al., 2004).

Training in the administration of complex instruments, such as intelligence tests, relies on test manuals and basic assessment texts, as well as on the clinical experience of supervisors. Test manuals vary considerably in the level of specific detail they prescribe and in the level of guidance they provide. Some tests, such as the Wechsler scales, demand strict adherence to detailed procedures for test administration. They state, for example, the specific distance a stimulus book must be placed from the edge of the table, the exact wording of examiner prompts, and the precise timings that are allowable for certain tasks. Other tests may be less prescribed, potentially producing ambiguity and uncertainty about aspects of test administration and scoring. Even if a high level of specific instruction is provided in the test manual, there will still likely be instances when examiners are unsure about how to deal with particular issues and dilemmas they encounter during testing. At these times, experienced clinicians may choose to consult a colleague, while interns inevitably depend on their supervisors.

It would be naïve to believe that psychologists who have been practising for many years know all the answers, or that they will necessarily agree with their similarly experienced colleagues on how to deal with particular assessment issues. For interns, such differences of professional

© Australian Psychological Society Ltd 2019.

CAMBRIDGE UNIVERSITY PRESS

opinion are likely to be confusing. Their current supervisor's clinical opinion may conflict with previous advice or training they have received. Although the supervisor's advice may be well-considered and based on substantial clinical experience, it could also potentially stem from well-practised errors that the intern may then proceed to perpetuate. Supervisors may be adhering to practices they learned in their own training and continue to apply automatically without sufficient critical reflection on their appropriateness. At times they may be unable to provide interns with adequate justification for these practices apart from 'this is the way I was taught'.

In this article we report the results of focus groups with intern educational and developmental psychologists. The aim of our investigation was to identify the assessment issues and dilemmas encountered by interns during their supervised practice; in particular, those aspects that are not manualised. The purpose was to provide data that could be used to improve training and supervised practice.

Method

Participants

The participants were nine postgraduate students (all female) who were enrolled in a master's program in educational and developmental psychology at an Australian university. The study took place at the end of the academic year, by which time all of the students had completed at least two first-year practicum units in the university clinic and/or in schools. Six of the participants were second-year students who had also undertaken two external practicums in workplaces such as schools, disability services, hospitals and community clinics. All students had been provided with structured training in Wechsler Intelligence Scale for Children (WISC-IV) test administration and scoring in the first semester of their master's program. This was followed by less intensive training on the Wechsler Individual Achievement Test (WIAT-II), the Woodcock Johnson battery, the Stanford Binet 5th Edition, and the Universal Nonverbal Intelligence Test (UNIT). In addition, they had been exposed to a range of other tests such as the Children's Memory Scale, the Leiter III, and the Test of Everyday Attention for Children (TEACh).

Measure

To identify the issues and dilemmas the interns faced during their practicum experiences, four stem questions were used in the focus groups: (1) What are some of the issues or dilemmas you have encountered while doing psycho-educational assessments? (2) Have there been any differences of opinion between one supervisor or lecturer and another with respect to certain aspects of assessment (e.g., choice of test, administration procedures, scoring, interpretation)? (3) Have any particular problems come up that were not covered in the test manual or in assessment texts? (4) Have there been times when you have felt frustrated, confused or anxious about your practice of psycho-educational assessment?

Procedure

Ethical clearance for the project was given by the University's Human Research Ethics Committee. Invitations to participate in the research were distributed to all 30 students who were currently enrolled in a master's program in educational and developmental psychology at Queensland University of Technology. If willing to be involved, students were asked to contact the research assistant who then organised and facilitated three focus group interviews. The sessions were conducted in a room on the university campus and lasted for around one hour. Each focus group was audio recorded and later transcribed.

Data Analysis

Thematic analysis was used to analyse the data following the guidelines proposed by Braun and Clarke (2006). The research assistant read the transcripts several times before identifying as many categories as possible. Initial codes were then created for the data. These codes were collated into themes and each response was placed into the most appropriate category. The authors then independently coded the responses and identified similar themes.

Results

Five main themes emerged from the focus groups: (1) Structure of the assessment session, (2) Strategies for encouraging children's participation, (3) Test administration, (4) Scoring and interpretation issues, (5) Professional issues.

Structure of the Assessment Session

Many of the interns expressed uncertainty about structural aspects of the assessment session, particularly with respect to breaks. They were unsure whether examinees should be told at the beginning of the session that they can ask for a break at any time, how many breaks are permissible, and for what reasons. One participant said she had decided that 'to get the best out of that child at a particular time — if they need a break, I give them a break' (Group 2, Speaker 1). A second issue was related to the kinds of activities that could occur during a break (e.g., is it acceptable to play a game or go for a walk) and, if a break is taken, how should testing be resumed (i.e., with what instructions/encouragement).

The interns mentioned receiving differing advice about splitting testing across two (or more) sessions: 'Some say you have to do it in one session and others say you can break it up into two sessions or do a couple of tests at a time if they're struggling' (Group 3, Speaker 2) with one participant mentioning an example in which a psychologist administered the WISC-IV across four separate sessions on four different days. There was uncertainty about the length of the gap that would be acceptable between two or more partial administrations, and which age norms to use if an examinee moved into a higher age bracket when the second part of the administration occurred.

Strategies for Encouraging Children's Participation

Interns were concerned about how to manage difficult child behaviour during testing. They were unsure what strategies could be used with resistant children, and many mentioned that behaviour potentially affected standardised administrations: 'I had one kid who refused to do anything and that was tricky and required a lot of encouragement and maybe an unstandardised approach to assessment' (Group 1, Speaker 1). There was a recognition that 'theory is very different from practice at times ... you want to follow the standardised manual but ... sometimes that just isn't possible' (Group 1, Speaker 1).

Other participants asked about the amount and type of encouragement, incentives and praise that can be given: '[it's hard] managing the social-emotional aspects, particularly when kids are struggling right in front of you, keeping their self-esteem high when they're struggling' (Group 2, Speaker 2).

https://doi.org/10.1017/edp.2019.3 Published online by Cambridge University Press

Test Administration

Although the participants acknowledged that strict adherence to standardised test administration had been emphasised during their training, several raised 'grey areas' that produced uncertainty. One such concern was about replacing American terms such as 'faucet' with Australian equivalents ('tap'). The use of prompts was also raised: 'You often end up saying more (than the standardised prompts) even though you're not supposed to' (Group 2, Speaker 3).

Many of the participants reported being surprised and confused when they observed supervisors and other psychologists who were not adhering strictly to standardised administration:

> The way we're taught the WISC, obviously we have to be very precise, but then when I've been on a placement and watched colleagues, they're not so precise ... that was a bit of a shock to see that the real world are using it very differently to the way we've been taught. (Group 3, Speaker 1)

Another said: 'We've been taught to make sure we say [the instructions] word for word, and they were just loosely saying instructions and loosely prompting, when we're told not to prompt in those sorts of ways' (Group 3, Speaker 2). Indeed, this mismatch between what the interns had been taught and what was happening in practice appeared to produce one of the most disturbing dilemmas: 'Staying standardised— how we are taught versus translating that into the real world is probably the biggest thing' (Group 2, Speaker 1).

Scoring and Interpretation Issues

Some participants reported difficulties with scoring verbal test items - in particular, the 'grey area between a (score of) 1 and a 2' (Group 2, Speaker 2). Whether to use age- or year-based scores when both options were available was a dilemma, with the interns reporting conflicting advice:

> My supervisor says you should always use age-based (on the WIAT) because that's what you've used on the WISC and if you want to be able to compare like that it just makes sense to be consistent. But ... other supervisors say ... it's about academic learning so it should always be year based ... it's sort of hard to know which way to go. (Group 2, Speaker 4)

Participants expressed confusion about reporting test results, particularly whether to give scores or confidence intervals or both: 'In schools they need the single scores, they're really big on that' (Group 3, Speaker 1). Another dilemma, described as 'huge' and 'very confusing' (Group 3, Speaker 2), was whether or not the Full Scale Intelligence Quotient (FSIQ) can be interpreted if there is variability among index scores. One participant said: 'Different supervisors seem to have different opinions, go on different criteria, like how big a discrepancy is and whether that matters or not' (Group 2, Speaker 4).

Report formats were said to 'depend on the supervisor' (Group 2, Speaker 3) because 'some supervisors like numbers and some like stories' (Group 2, Speaker 4). The interns were sometimes told: 'Just give me the numbers' (Group 2, Speaker 5).

Professional Issues

The participants raised several professional issues related to time pressures, lack of adequate resources, and disagreements with employers or supervisors. Time pressures meant that some tests had to be learned without supervisor input: 'You learn it by yourself ... there's no time to practise with your supervisor to see if you're doing it right' (Group 3, Speaker 2), and at times supervisors appeared not to have 'a lot of experience in administering it either' (Group 2, Speaker 5).

Sometimes in external practicum settings the most appropriate tests were not available, and several participants mentioned having to use whatever tests were available 'even though you know it's not right' (group 2, speaker 3). One gave the example of being told she had to use a Wechsler test for a student with profound speech and language impairments because there was no access to a non-verbal test of intelligence. Another reported finding it difficult when schools wanted IQ scores even when these were clearly inappropriate:

> I do find it tricky to explain to schools that it's not really appropriate ... the child has speech and language delays and fine motor difficulties and you know that they're going to come out really low [on an IQ test]. (Group 1, Speaker 1)

In particular, disagreements were noted with respect to diagnosis. One intern said that 'diagnosing or not diagnosing (depends) on the organisation or who your supervisor is' (Group 2, Speaker 2). The students reported being unable to use dyslexia as a diagnosis because '[the school] doesn't like you using dyslexia' (Group 1, Speaker 1). It was reported to be particularly difficult when supervisors 'said that dyslexia doesn't exist ... I found that hard ... learning a lot about it and reading the research, and then it being a commonly held view across the supervisors that it doesn't exist' (Group 3, Speaker 2).

Disagreements with supervisors were particularly stressful for the interns. Some reported following their supervisor's advice — 'I've just sort of trusted their judgment' (Group 2, Speaker 5) — even when it conflicted with what they believed to be correct: 'At the end of the day you have to go with your supervisor ... and you just feel that it's wrong, it's wrong, it's wrong' (Group 2, Speaker 3). One participant said she tried to present her own opinion as well: 'I've ended up mostly going with what my supervisor has said, but making a point of what I think as well' (Group 2, Speaker 4), and another mentioned an instance when she justified her decision on a particular issue: 'I was able to find it in the book and show them and say: "Look, I was on the right track"' (Group 3, Speaker 1). With respect to reconciling differences of opinion, participants seemed to favour their training over their supervisor: 'I think we'd go with the way we've been taught because you're safe then' (Group 2, Speaker 2).

Although conflicting views were usually discussed in supervision, an instance was mentioned of a supervisor making significant changes to an intern's report without any discussion:

> She [the intern] had concluded that [the child] wasn't ready for prep and needed more time at ECDP [the Early Childhood Development Program] and the supervisor edited it and said he was ready for prep and the student [intern] wasn't comfortable with that. (Group 1, Speaker 1)

Discussion

This research provides insights into the issues and dilemmas that provisionally registered educational and developmental psychologists encounter in their practice of psycho-educational assessment. Most of their concerns relate to experiences in the 'real world' where they faced situations that their basic level of training had not prepared them for and that were not covered in test manuals or textbooks.

Many of the interns mentioned uncertainties about structural aspects of testing such as breaks and incentives for examinees. Most test manuals and assessment texts state that breaks are

https://doi.org/10.1017/edp.2019.3 Published online by Cambridge University Press

permissible, although their duration and content depend on the judgment of individual psychologists. Splitting a single test across multiple days is undesirable, although admittedly sometimes necessary, especially for children who work slowly or fatigue easily (Sattler, 2018). However, testing should not extend beyond two sessions, preferably no more than a week apart, and some cautions are needed for interpreting results whenever there are deviations from procedures used in the norming sample.

Opinions about the use of praise and incentives for encouraging children's participation in testing vary in the published literature. Encouraging comments such as 'It's OK to guess even if you're not sure' appear to be universally acceptable unless specifically disallowed in a test manual. Examiner comments that recognise effort can be especially motivating but need to be provided judiciously. However, the use of tangible, especially edible, incentives is more controversial and usually recommended only in exceptional circumstances (Kranzler & Floyd, 2013; Sattler, 2018).

Frustration about being unable to adhere to best practice was common in the current study. A major issue for interns was accessing the most appropriate tests, finding time to become competent in their administration, and obtaining the necessary supervision to ensure competency. The participants were well aware that these difficulties potentially resulted in unacceptable assessment practices. Psychologists have the ultimate responsibility for evaluating a test's appropriateness for a specific client and referral question (Conoley, Conoley, & Hernandez, 2013), but convincing employers such as school systems and community agencies of the need to purchase tests beyond those with which they are most familiar (e.g., the Wechsler scales) can be challenging.

It is clear that interns experience much confusion and frustration when supervisors have differing views, or when a supervisor's advice is inconsistent with concurrent university training. Stoltenberg and McNeill (2012) reported that differences in theory and practice are a major cause of tension in the supervisee-supervisor relationship. Some supervisors need to adapt their assessment practices to match the demands of the organisations in which they work (Dodds, 1986), and this could be especially true within school systems where the participants in our study cited examples of disagreements about reporting IQ scores and diagnosing dyslexia.

Many of the issues and dilemmas identified by the interns are not unique to them but are commonly shared by all psychologists, irrespective of experience. In the absence of clear guidelines, psychologists make decisions based on their training, experience and clinical judgment. Some decisions may be about seemingly inconsequential aspects of psycho-educational assessment, such as whether test record forms should be completed in pencil or pen. This choice may seem trivial at first glance; however, there are practicalities to be considered, as test record forms usually need to be retained for more than 20 years. Pencil does not fade over time as readily as ink, but can be easily erased and changed at a later time.

It is important for psychologists to be able to explain and justify the decisions they make in their practice of psychological testing (American Educational Research Association [AREA], American Psychological Association & National Council on Measurement in Education, 2014). For example, what is the justification for using the 95th versus the 90th confidence interval when both options are available? Or what is the justification for selecting a specific index score discrepancy to rule out the use of the FSIQ? With respect to the former question, most sources appear to recommend using the 95th confidence band (e.g., Harvey, 2013) although usually without a specific justification for this choice. Kranzler and Floyd (2013) suggest that the 90th confidence interval is easier to describe (i.e., a probability of 9 out of 10). In relation to the latter question, recent publications have suggested that the FSIQ is almost always interpretable and reportable (see Watkins, Dombrowski, & Canivez, 2018). Of concern is the possibility that decisions about issues such as these are sometimes made automatically and unthinkingly without sufficient reflection, peer discussion or consultation with the current literature. Yet the various professional, ethical and legal standards and guidelines for psychologists provide clear expectations for responsible professional practice, with reflection and critical self-evaluation being essential components (AERA et al., 2014).

Although the current study involved only a small sample of intern psychologists at one university, the findings provide insights into the issues and dilemmas that need to be explored in the teaching of psycho-educational assessment by both universities and workplace supervisors. One of the major issues we identified was the deviation from standardised test administration that interns observed in workplaces. Such deviations appear not to be unusual among experienced psychologists (Wolfe-Christensen & Callahan, 2008). The risk is that interns may conclude that their university training involved unrealistically high standards that can now be lowered in the 'real world'. The second major issue was the differing advice interns received from different supervisors. Although the power imbalance in favour of supervisors potentially increases the likelihood that interns will change their practices in line with a supervisor's advice (albeit perhaps only temporarily), some of the participants in this study appeared willing to defend the decisions they made on the basis of their training.

Clearly, it would be beneficial for both interns and experienced psychologists to be able to draw on research evidence and professional guidelines with respect to issues and dilemmas such as those raised here. Unfortunately, such evidence appears to be almost non-existent and practice guidelines are often elusive. Expert opinions vary, and there are few accepted rules for best practice beyond the basics that are covered in test manuals and assessment texts (e.g., Kranzler & Floyd, 2013; Sattler, 2018) or the more complex contemporary issues that are discussed in advanced publications (e.g., Decker, Hale, & Flanagan, 2013; Wilcox & Schroeder, 2015). Irrespective of the availability of appropriate research evidence and guidelines, we urge trainers and supervisors to instil in their students the absolute importance of reflective and critical assessment practice throughout their professional careers.

Acknowledgments. The authors would like to thank the postgraduate students who so willingly and enthusiastically volunteered to participate in this research.

Financial Support. The research was supported by a Queensland University of Technology Learning and Teaching Grant to the authors.

Conflicts of Interest. None.

Ethical Standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

American Educational Research Association (AREA), American Psychological Association & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, DC: AERA.

https://doi.org/10.1017/edp.2019.3 Published online by Cambridge University Press

Belk, M.S., LoBello, S.G., Ray, G.E., & Zachar, P. (2002). WISC-III administration, clerical, and scoring errors made by student examiners. *Journal of Psychoeducational Assessment*, *20*, 290–300.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Butcher, J.N. (2006). Assessment in clinical psychology: A perspective on the past, present challenges, and future prospects. *Clinical Psychology: Science and Practice*, *13*, 205–209.

Conoley, J.C., Conoley, C.W., & Hernandez, R.J.C. (2013). Technical reporting, documentation, and evaluation of tests. In K.F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 251–263). Washington, DC: American Psychological Association.

Decker, S.L., Hale, J.B., & Flanagan, D.P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools*, *50*, 300–313.

Dodds, J.B. (1986) Supervision of psychology trainees in field placements. *Professional Psychology: Research and Practice*, *17*, 296–300.

Gilmore, L., & Campbell, M. (2009). Competency in intelligence testing: A training model for postgraduate students. *Australian Educational and Developmental Psychologist*, *26*, 165–173.

Harvey, V.S. (2013). Communicating test results. In K.F. Geisinger, B.A. Bracken, J.F. Carlson, J.C. Hansen, N.R. Kuncel, S.P. Reise, and M.C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology: Vol. 2. Testing and assessment in clinical and counseling psychology* (pp. 35–50). Washington, DC: American Psychological Association.

Kranzler, J.H., & Floyd, R.G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. New York, NY: Guilford Press.

Krishnamurthy, R., VandeCreek, L., Kaslow, N.J., Tazeau, Y.N., Miville, M.L., Benton, S.A. (2004). Achieving competency in psychology assessment: Directions for education and training. *Journal of Clinical Psychology*, *60*, 725–739.

Sattler, J. (2018). *Assessment of children: Cognitive foundations* (6th ed.). San Diego, CA: Author.

Stoltenberg, C.D., & McNeill, B.W. (2012). Supervision: Research, models, and competence. In N.A. Fouad (Ed.), *APA handbook of counseling psychology: Vol. 1. Theories, research and methods* (pp. 295–328). Washington, DC: American Psychological Association.

Watkins, M.W., Dombrowski, S.C., & Canivez, G.L. (2018). Reliability and factor validity of the Canadian Wechsler Intelligence Scale for Children — Fifth edition. *International Journal of School & Educational Psychology*, *6*, 252–265.

Wilcox, G., & Schroeder, M. (2015). What comes before report writing? Attending to clinical reasoning and thinking errors in school psychology. *Journal of Psychoeducational Assessment*, *33*, 652–661.

Wolfe-Christensen, C., & Callahan, J.L. (2008). Current state of standardization adherence: A reflection of competency in psychological assessment. *Training and Education in Professional Psychology*, *2*, 111–116.

https://doi.org/10.1017/edp.2019.3 Published online by Cambridge University Press