

Constructing the Human Figure Drawing Continuum: One Scale is ‘Good Enough’

Claire Campbell and Trevor Bond

Introduction

Historically, drawing has been considered an innate form of expression for humans (Fowlkes 1980). Prior to the development of verbal communication systems, our pre-historic ancestors used drawings to convey information. Indeed, the first crude form of written language—known as ‘proto-writing’—consisted of tiny, intricate drawings rather than early forms of letters or numerals (Houston 2004). Pre-historic cave drawings and present-day young children’s drawings could be considered to share a similar underpinning motive; both used drawing to convey information, understanding and knowledge during a time when other forms of communication were not yet fully mastered (Fowlkes 1980). Indeed, today’s young children characteristically delight in drawing representations of what they know and understand about the world, while still acquiring verbal and written communication skills (Di Leo 1970). Whilst the notion of children’s ‘drawings’ typically evokes discussions about art, creativity and imagination, drawings reveal much more developmental and intellectual content than most realise.

Children’s drawings have been researched from a variety of perspectives for over 150 years. Children’s drawings of the human figure are amongst the most researched (Cox 1992; Kellogg 1967, 1970; Koppitz 1968). This is not surprising to many, as young children’s first attempts at creating a human figure drawing (HFD)—known as cephalopods (see Fig. 1)—are amongst the most distinctive and recognisable aspects of early childhood. The wealth of research indicates that there is a general developmental sequence to children’s HFDs (Cox 1992, 1997; Di Leo 1970, 1973; Luquet 1913; Maley 2009; Mavers 2011; Piaget and Inhelder 1956, 1971). That is, similar to how children typically crawl, walk and then run sequentially in the domain of gross motor development (Goodway et al. 2012),

C. Campbell (✉) · T. Bond

College of Arts, Society and Education, James Cook University, Townsville, Australia
e-mail: claire.campbell2@jcu.edu.au

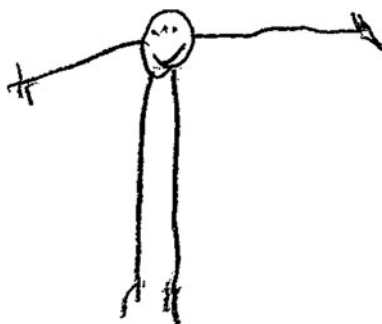


Fig. 1 Example of a cephalopod drawing by a young child

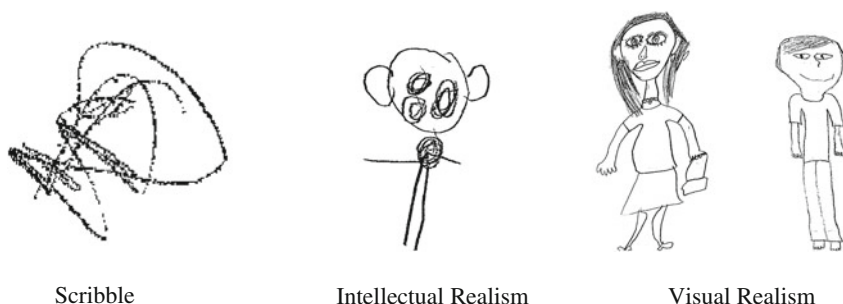


Fig. 2 Examples of children's drawings at three stages (Piaget and Inhelder 1956, 1971)

children sequentially make scribbles in the 'fortuitous realism' or 'scribbling' stage, and then 'projective' and 'imaginal representations' of humans in the 'intellectual realism' stage, followed by drawings with more accurate proportions and perspective in the 'visual realism' stage in the area of HFD development (see Fig. 2) (Piaget and Inhelder 1956, 1971).

Piagetian theory provides a useful lens for investigating young children's drawings. Jean Piaget (1896–1980) was the first to systematically study children's cognitive development and the findings of his Genevan school have had significant impact on the fields of education and psychology. Piagetian theory—which describes children's cognitive development as progressing through an invariant sequence of age-related (not age-dependent) stages of thinking: sensorimotor (birth to toddlers); preoperational (early years education); concrete operational (primary school) and formal operational (secondary school and onwards)—is well known in educational and developmental psychology circles. However, many parents, carers of young children and lay people remain unfamiliar with his theories. Consequently, when children display behaviours that could be considered characteristic of the preoperational stage of thought—such drawing of cephalopods—many people dismiss this as just some of the 'bizarre' things about early childhood. Indeed, many

parents and carers might attempt to teach young children to include necks, torsos, shoulders and similar features in their HFDs or, in a related example, to explain to them that they do not have more cake to eat now that it has been cut into pieces as opposed to its being left whole. However, adherents to Piagetian theory would suggest that these efforts would be to no avail; until the child develops more organised mental structures, the benefit of this sort of didactic teaching cannot be realised (Piaget 1971).

Piaget and Inhelder (1956, 1971) examined young children's 'representations', or drawings, as an extension of their earlier work on cognitive development. In brief, their theory suggests that the 'borderlines' between the child's perception of an object, the child's mental representation of that object and the child's actual pictorial representation of that object create *discontinuities* in the drawing process. As children's thinking progress through the developmental stages, become increasingly more able to transition with *continuity* through these borderlines and, consequently, produce more comprehensive and realistic pictorial representations (Piaget and Inhelder 1956, 1971). As young children's thinking transitions toward the concrete operational stage of thought, their perception becomes progressively decentred; they are more able to integrate information from a variety of sources and reorganise their knowledge into more sophisticated systems. As a result, older and more experienced children become increasingly able to see others' points of view, consider multiple aspects of a problem at one time and produce more logical and comprehensive explanations and drawings.

Many researchers agree on the links between young children's HFDs and their levels of cognitive development and understanding (Anning and Ring 2004; Cox 1992; Di Leo 1973; Goodenough 1926; Harris 1963; Piaget and Inhelder 1956, 1971). The most influential work in this body of knowledge is held to be that originally undertaken by Goodenough (1926) and Harris (1963).

Background

Development of the Goodenough Draw-a-Man Test (GDAMT)

Florence Goodenough (1886–1959) was an American teacher interested in intelligence and conventional intelligence (IQ) tests (Goodenough 1949; Goodenough and Tyler 1959). She worked under the supervision of Lewis Terman at Stanford University developing the Stanford-Binet Intelligence Quotient test for children (Thompson 1990). Her doctoral research (Goodenough 1924) into young children's human figure drawings resulted in the GDAMT, the world's first non-verbal assessment for inferring children's levels of cognitive development and understanding via the details included in their drawings of an adult male human figure. Her published doctoral thesis, *Measurement of Intelligence by Drawings*

(Goodenough 1926), detailed the development of the 51-item test and explained the strategic selection of an adult ‘male human’ as subject matter. Goodenough’s research indicated that other drawings, including those of women, houses, animals and the like, were not sufficiently ‘uniform’ and thereby produced erratic results using the statistical methods of the time (Goodenough 1926). Unlike the other non-verbal assessments available at that time, the GDAMT exhibited high reliability and validity, was easy to administer, and correlated well with other intelligence tests (Goodenough 1926; Harris 1963).

Development of the Goodenough-Harris Drawing Test (GHDT)

Goodenough’s doctoral student, Dale Harris (1915–2007), revised and extended the original GDAMT; it is now known as the GHDT. Harris made two key changes: (1) he extended the original 51-item GDAMT to a 73-item Draw-a-Man (DAM) sub-test; and (2) he included a 71-item Draw-a-Woman (DAW) sub-test and a Self-Portrait sub-test (73-item SPM sub-test for male children and 71-item SPF sub-test for female children). As with the original GDAMT, the GHDT yielded high reliability and validity and correlated well with the other more conventional intelligence tests using the statistical methods available at the time (Harris 1963). Notwithstanding, Harris added the DAW and SPM/SPF sub-tests even in the face of a complete lack of empirical evidence indicating that a single drawing of a man was actually insufficient for the assessment task. Furthermore, Harris did not verify empirically the effectiveness of the tripled data collection load (i.e., the collection of three drawings instead of one) for each child.

Research Objective

This study adopted the Rasch model for measurement (Rasch 1960) to (1) examine the psychometric properties of the GHDT and young children’s HFDs, and (2) investigate what additional information might be revealed by Harris’s additional items for the DAM sub-test, as well as the DAW and SP sub-tests.

Method

Sample

A sample of 107 children of different ages and abilities (aged between 4 and 10 years) were recruited from a Preparatory to Year 12 school in Queensland,

Australia (Preparatory, or 'Prep', is the year level prior to the first year of formal education). A total of 738 HFDs were collected from these children over three phases of data collection, approximately six months apart. The sample included children from a range of socioeconomic backgrounds, however, the ethnic backgrounds of the participants were not as diverse. Most of the sample was of Caucasian descent with approximately 3 % of children from Australian Aboriginal, Torres Strait Islander and Asian backgrounds. The children's HFDs were not examined according to socioeconomic status or ethnicity. All participants had informed parental consent to participate in the study.

Data Collection

The children's HFDs were collected via the published GHDT administration procedures, which simply require children (either individually or in small groups) to make three drawings, one each of a man, a woman and a self-portrait, to the best of their ability. Provided that children attempt to draw a whole person (that is, not a 'bust'-type portrait), there is no 'right' or 'wrong' type of drawing. During the small group data collection procedures, children were seated so that they could not 'copy' from another's drawings. The GHDT administration does not have a time limit, although most children complete all three drawings in approximately 10–15 minutes. All HFDs were examined and scored (by the first author) in alignment with the two GHDT scoring guides: a 73-item guide for the DAM sub-test and a 71-item guide for the DAW sub-test. Whilst the scoring guides share 50 common items, most item numbers do not correspond across the guides. Moreover, the SPM and SPF sub-tests do not have dedicated scoring guides; they are scored merely using the DAM or DAW scoring guide as appropriate to the reported gender of the child.

Data Analysis

The qualitative drawings were converted into quantitative data so that the Rasch model could be applied. This was quite a straightforward process as the scoring guides used a dichotomous (0, 1) system for each drawing criterion. That is, if the item was judged as absent from the HFD or present but not in fulfilment of the criterion, zero (0) credit was assigned to that item number. Conversely, if the item was present in the HFD and satisfying the criterion, one (1) credit was assigned to that item number. Once scoring was complete, data files were created using Excel[®] software and then submitted to WINSTEPS[®] version 3.68.0 (Linacre 2009) for Rasch analysis.

Results

Overview

Rasch analysis of the data produced a range of output including variable maps and summary statistics. The results of the Rasch analysis of the DAM, DAW, SPM and SPF sub-tests revealed quite remarkable similarity. The equal-interval logit measurement scales produced by the Rasch analysis of the DAM and DAW sub-tests each spanned a range of 14 logits (from -8 to $+6$ logits). The SPM and SPF logit scales were only two logits shorter spanning from -7 to $+5$. Interestingly, most of the 50 test items common to all four sub-tests (e.g., head, eyes, trunk/body, legs, nose 2D) were located in similar positions along the logit scales across all four variables maps for the DAM, DAW, SPM and SPF sub-tests.

A selection of key summary statistics produced by the Rasch analyses of the GHDT sub-test data is presented together in Table 1. The total number of items and persons, means, standard deviations, reliability, separation and number of underfitting and overfitting items and persons are displayed for the each of the four GHDT sub-tests, respectively.

Unidimensionality

The comprehensive work completed by Goodenough (1926) and Harris (1963) was completed using the analytical methods appropriate for the time. Subsequently,

Table 1 Selected Rasch analyses summary statistics for the four GHDT sub-tests

	DAM	DAW	SPM	SPF
Items				
N	73	71	73	71
Mean	0.00	0.00	0.00	0.00
SD	2.90	2.77	2.48	2.71
R	0.98	0.99	0.97	0.98
Separation	7.89	8.37	5.48	6.72
Underfitting (n)	10	6	5	12
Overfitting (n)	9	12	2	7
Persons				
N	246	246	99	147
Mean	-1.26	-1.18	-1.22	-0.88
SD	1.45	1.43	1.3	1.41
R	0.93	0.92	0.89	0.93
Separation	3.55	3.50	2.89	3.54
Underfitting (n)	10	6	5	4
Overfitting (n)	2	4	0	5

neither young children's HFDs, nor the GHDT, have been investigated comprehensively from a modern test theory perspective. Correspondingly, thus far it has presumed that all items in the GHDT sub-tests contribute meaningfully to the investigation of a single underlying drawing construct. The Rasch model's unidimensionality principle requires that all items comprising the GHDT investigate only one construct at a time—in this case, the development of children's HFDs. The Rasch model's fit statistics and other diagnostic indicators helped to determine that most items and persons under investigation satisfied this unidimensionality requirement. Rasch analyses of the data indicated only a small number of erratic (i.e., underfitting) items for each of the GHDT sub-tests (see Table 1). Similarly, even fewer items were detected as being overly predictable (i.e., overfitting) across the four sub-tests (see Table 1). The person fit statistics revealed a very small number of 'misfitting' person measures across each of the GHDT sub-tests; however, there were no children who consistently presented misfitting performances.

GHDT Sub-test Comparisons

Given that the GHDT did not breach the Rasch model's unidimensionality principle, the second step in this study was to investigate what additional information, if any, was revealed by Harris's additional DAW and SPM/SPF sub-tests. Closer inspection of the person fit statistics and ability estimates revealed that children performed almost identically on all three GHDT sub-tests that they completed. Consequently, very little additional information is revealed by the DAW and the SP sub-tests over that already revealed by the DAM sub-test. Given that mean person estimates close to zero (0) are indicative of a well-matched test (Bond and Fox 2015), the SPM (0.42) and SPF (0.38) sub-tests were found to be better targeted to this sample of children than were the DAM (−1.26) and DAW (−1.18) sub-tests. Despite this, children tended to receive slightly lower raw scores for their SP drawings than they did for their drawings of adult males and females. This could be linked to the fact that the SP sub-tests are scored using the applicable adult 73-item DAM or 71-item DAW scoring guides. These guides were designed for scoring drawings of adults and contain item-scoring criteria that should be considered problematic for young children's drawings of themselves. In particular, the DAW scoring guide contains a number of items pertaining to jewellery, clothing, high heel shoes, breasts, hips and the like, which—understandably—might prove problematic for several developmental, cultural and societal reasons. A young girl's self-portrait might not include several (or all) of these items, as they are simply irrelevant developmentally, culturally and/or socially. However, the scoring of such a SPF drawing using the DAW scoring guide assumes that these items are 'absent' due to a lack of understanding, rather than inappropriateness due to the drawer's level of physical development, cultural circumstances and/or social background. It

was issues such as these that prompted the development of a modified GHDT that contained only items applicable to drawings of all human figures—men, women, boys and girls.

Development of the Human Figure Drawing Continuum

Eventually, there were several key factors that pointed towards the need for a more parsimonious and developmentally/culturally/socially relevant human figure drawing test: (1) the finding that similar items across all four sub-tests were plotted in similar locations along the logit scale in the person-item maps; (2) the finding that little additional information is yielded from the two additional sub-tests beyond that already revealed by a single drawing and (3) the finding that some items in the scoring guides were irrelevant or inappropriate to the scoring of drawings of children. Given that there were 50 common items across the DAM/SPM and DAW/SPF sub-test scoring guides, a key analytical task was to investigate whether all or any particular combination of those 50 (or fewer) common items could be used to measure effectively *any* HFD made by young children.

The development of the Human Figure Drawing Continuum (HFDC) commenced with an examination of the fit statistics for each of the 50 common items across the four GHDT sub-tests. The 50 common items were deemed to form a sound ‘starting point’ as each of them is applicable to drawings of males and females, regardless of age. After careful examination of the fit statistics, various combinations of misfitting items were removed iteratively in an effort to enhance the measurement properties of the proposed instrument, and repeated Rasch analyses were conducted. Each iteration produced output that was inspected closely for adherence to Rasch’s measurement expectations until finally a particular combination of 45 items was identified as best fitting the Rasch model’s requirements. That is, additional items or fewer items did not improve the person and item fit statistics, variance explained, standard deviations, person separation indices or person and item reliability values.

Common person linking, or invariance, graphs (Bond and Fox 2015) were used to verify the effectiveness of the HFDC as a scoring mechanism in comparison to each of the four GHDT sub-tests. Figure 3 displays the DAM *v.* HFDC and the SPM *v.* HFDC and Fig. 4 shows the DAW *v.* HFDC and the SPF *v.* HFDC, respectively. These graphs of the paired person measures from the 45-item HFDC against each of the 70-plus item GHDT sub-tests indicate that all comparisons yield congruent results. That is, all graphs show all plots located between the 95 % control lines indicating the invariance of the person measures within error—except for one, single, anomalous performance. This male child aged seven years was ‘on the borderline’ in the DAW *v.* HFDC common person linking graph (Fig. 4) with a DAW estimate of -2.66 logits (err = 0.44) and a HFDC estimate of -1.43 logits (err = 0.49). That is, he drew a ‘human’ more successfully than he drew a ‘woman’.

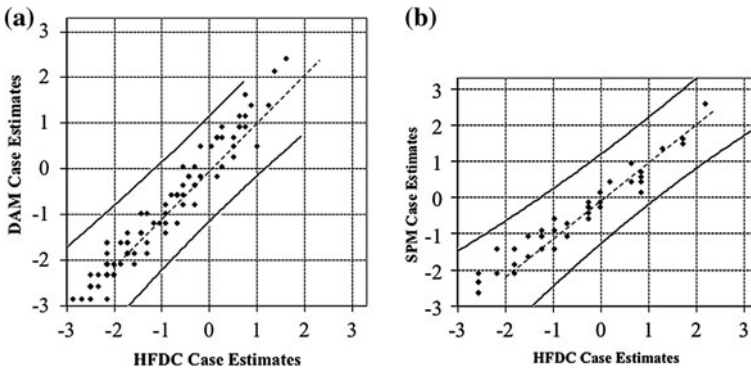


Fig. 3 Common person linking graphs: **a** 73-item DAM v. 45-item HFDC and **b** 73-item SPM v. HFDC

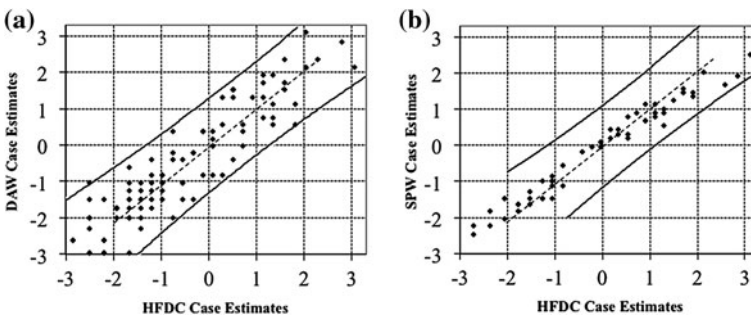


Fig. 4 Common person linking graphs: **a** 71-item DAW v. 45-item HFDC and **b** 71-item SPF v. HFDC

The graphs (Figs. 3 and 4) clearly indicate that the 45-item HFDC was almost exactly as effective as any of the 70-plus item GHDT sub-tests in assessing young children's HFDs. Indeed, it could be argued that the HFDC was the best suited and most reliable of all of the instruments. First, the HFDC yielded an item standard deviation of 3.04 indicating that it had the greatest spread of items along the logit scale in comparison to the DAM (2.90), DAW (2.77), SPM (2.48) and SPF (2.71) sub-tests. Second, the HFDC yielded an item reliability of 0.99 (i.e., congruent with the other sub-tests: DAM: 0.98; DAW: 0.99; SPM: 0.97; SPF: 0.98) suggesting that the item hierarchies are equally likely to be replicated when the instrument is administered to other suitable samples. Third, the HFDC yielded a person reliability index of 0.89, which is not meaningfully different from that produced by the DAM, DAW and SPF sub-tests (0.92) and the SPM sub-test (0.88). Fourth, the HFDC yielded the highest mean person measure (-0.59) of all the sub-tests (DAM: -1.26 ; DAW: -1.18 ; SPM: -1.22 ; SPF: -0.88) indicating that it is better targeted to the abilities of this sample of children. And, finally, whilst the HFDC person separation

the left-hand side of the logit scale indicates that three children were located at that estimated ability level ('.' indicates $0 < n < 3$). At 14 logits, the logit scale range is just as expansive as those produced from the DAM and DAW sub-test data. Some item numbers have been replaced with the item name to help add substantive meaning to the HFDC scale.

Discussion

This study was the first to investigate the GHDT and young children's HFDs from a modern test theory perspective. The principle aims were to (1) examine the psychometric properties of the GHDT and young children's HFDs and (2) investigate the potential redundancy of Harris's additional sub-tests. Given that the GHDT and the HFDs were deemed to be psychometrically sound—and that Harris's revision and extension revealed little beyond that revealed by a single drawing—the possibility of a more parsimonious instrument was explored. Therefore, 45-item instrument HFDC was identified and verified to be just as effective as any of the four 70-plus item GHDT sub-tests.

Implications

This research has shown that, despite Harris's good intentions, children's separate drawings of men, women and themselves do not provide additional information that is useful to the test administrator. Furthermore, for this assessment, children's HFDs do not need to be scored against 70-plus criteria. A single drawing of a self-selected human figure (man, woman or child) scored against a 45-item scoring guide should reveal almost exactly the same information as three drawings and some 200-plus drawing and scoring criteria. The authors have not been able to locate information on why Harris decided to revise and extend the original GDAMT; however, it could be assumed that it was linked to societal pressure regarding inclusivity, ageism and sexism. This research has revealed that the assessment can be made more inclusive and non-sexist by merely asking children to self-select a single human figure drawing rather than by requiring children to draw one each of a man, a woman and a child.

Another implication of this research is the empirical evidence as to which HFD concepts are least and most difficult for children to include. The DAM, DAW, SPM, SPF and the HFDC all had extremely similar item hierarchies further confirming, via modern test theory, the notion that young children's HFDs are underpinned by a common developmental sequence.

Limitations

Whilst the proposed HFDC mitigates many of the limitations of the GHDT, like all research, this study has its own limitations. Foremost, the HFDC lacks the wealth of empirical support of both the GDAMT and the GHDT. Future research should involve replication to investigate whether similar results can be achieved with a larger sample of more diverse children. Also, ideally, a project could follow the development of HFDs from the very first marks produced at around 18 months of age through adulthood to investigate the HFD developmental process. Further, this research investigated only the GHDT from a modern test theory perspective. It did not correlate or co-calibrate the GHDT, or the HFDC, with other assessments (such as intelligence tests). As all data were collected from one school in Queensland, Australia, appropriate caution must be applied when transferring findings to other settings. Last, this research did not investigate the links between children's HFDs and their levels of cognitive development; it relied on the research already undertaken by Goodenough and Harris in that regard. Moreover, the data were not collected as representing the proposed HFDC, *a priori*; this should form a key feature of subsequent investigations. Future research could investigate young children's HFDs and their levels of cognitive development—via the administration of Piagetian conservation tasks, for example—to examine what relationship exists between the two.

Conclusion

Given that the original GDAMT was created in 1926, and Harris's revision and extension was completed in 1963, it could be concluded that the drawing assessment was due for re-investigation, particularly from a modern test theory perspective. This research confirmed that the GHDT and young children's HFDs are both apt for Rasch analysis, and that they both satisfied the Rasch model's unidimensionality principle and its other measurement requirements. Interestingly, the application of the Rasch model to the GHDT has taken us *back* to what Florence Goodenough found in the first instance—that a single human figure drawing scored according to 50 or so criteria is 'good enough'.

References

- Anning, A., & Ring, K. (2004). *Making sense of children's drawings*. Berkshire, England: Open University Press.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). United Kingdom: Routledge.
- Cox, M. (1992). *Children's drawings*. London: Penguin Group.

- Cox, M. (1997). *Drawings of people by the under-5s*. London: Falmer Press.
- Di Leo, J. H. (1970). *Young children and their drawings*. New York: Brunner/Mazel Inc.
- Di Leo, J. H. (1973). *Children's drawings as diagnostic aids*. New York: Brunner/Mazel Inc.
- Fowlkes, S. S. (1980). *The effect of colour on human figure drawings as related to level of social adaptability*. Unpublished Master thesis. United States of America: The Ohio State University.
- Fisher, W. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions*, 6(3), 238.
- Goodenough, F. L. (1924). *The intellectual factor in children's drawings*. Doctoral thesis. Stanford University.
- Goodenough, F. L. (1926). *The measurement of intelligence by drawing*. New York: World Book Co.
- Goodenough, F. L. (1949). *Mental testing: Its history, principles and applications*. New York: Rinehart and Company Inc.
- Goodenough, F. L., & Tyler, L. E. (1959). *Developmental psychology* (3rd ed.). New York: Appleton-Century-Crofts Inc.
- Goodway, J., Ozmun, J., & Galladue, D. (2012). Motor development in young children. In O. N. Saracho & B. Spodek (Eds.), *Handbook of research on the education of young children* (3rd ed.). United Kingdom: Routledge. Retrieved from http://search.credoreference.com.elibrary.jcu.edu.au/content/entry/routsmch/motor_development_in_young_children/0.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough draw-a-man test*. New York: Harcourt, Brace & World Inc.
- Houston, S. D. (Ed.). (2004). *The first writing: Script invention as history and process*. Cambridge: Cambridge University Press.
- Kellogg, R. (1967). *The psychology of children's art*. San Diego: Random House.
- Kellogg, R. (1970). *Analysing children's art*. Pal Alto, California: Mayfield.
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune & Stratton.
- Linacre, J. M. (2009). *WINSTEPS*® Version 3.68.1 [computer software]. Chicago, IL: <http://www.winsteps.com>.
- Luquet, G. H. (1913). *Les dessins dun enfant*. Paris: Librairie Felix Alcan.
- Maley, C. (2009). *Young children's human figure drawings: An investigation using the Goodenough-Harris drawing test and the Rasch model for measurement*. Unpublished Ph.D. thesis. Australia: James Cook University.
- Mavers, D. (2011). *Children's drawing and writing: The remarkable in the unremarkable*. New York: Routledge.
- Piaget, J. (1971). *Biology and knowledge*. Chicago: University of Chicago Press.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge and Kegan Paul.
- Piaget, J., & Inhelder, B. (1971). *Mental imagery in the child*. London: Routledge & Kegan Paul Ltd.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Thompson, D. (1990). Florence Laura Goodenough. In A. O'Connell & N. Russo (Eds.), *Women in psychology: A bio-bibliographic sourcebook*. Westport, CT: Greenwood Press.

Pacific Rim Objective Measurement Symposium (PROMS)

2015 Conference Proceedings

Zhang, Q. (Ed.)

2016, XXIII, 424 p. 72 illus., Hardcover

ISBN: 978-981-10-1686-8