Paper code: 1728

The validity of constructing a measure of school need for psychological services

Adrian Young
Professor Rob Cavanagh
Curtin University of Technology
adrian44@iinet.net.au

Abstract
The construction of a new instrument ought to proceed through a series of sequential stages. Careful documentation of the developing process forms an important and necessary step in formulating a validity argument as evidence to support applications of the new measure. Furthermore, this process has crucial implications for the credibility of any inferences that may be drawn from applying the instrument as an intended measure. Recognising examples of evidence presented the opportunity to appraise the usefulness of the Wolfe and Smith (2007a & 2007b) validity framework which identified seven aspects of validity evidence.

The paper presents a summary of the construction of a measure of school need for psychological services and data analyses. Then the activities and findings were assessed as examples of validity evidence with respect to the Wolfe and Smith (2007a & 2007b) framework.

In addition, publicly available school-level data were collected such as socio-economic index, suspension and exclusion data, truancy and students with individual behaviour management plans. These were treated as independent variables and the linear measure of school need for psychological services was treated as the dependent variable. Hypothesised associations between the dependent variable and the independent variables were tested by multiple regression analysis of data from 148 school staff members. Each of the school-level indices and the dependent variable were also examined by an estimation of bivariate correlation.

The study identified examples of six aspects of validity evidence in the empirical investigation of school need for psychological services. This confirmed the usefulness of the multi-level theory of validity evidence postulated by Wolfe and Smith (2007a & 2007b). In addition, the analyses and graphical displays generated by the RUMM2020 computer program (Andrich, Sheridan, Lyne & Luo, 2003) proved invaluable in illustrating validity evidence. The measure of school need for psychological services was found to be significantly related to student suspensions data. As a result, application of the measure can inform decisions about the level of psychological services that should be provided to schools congruent with the psychological needs of their students.

The validity of constructing a measure of school need for psychological services.

Introduction
The paper begins by examining the necessity for school psychology service delivery to adapt to the rapidly changing nature of schooling. The absence of systemic attempts to document the functions of school psychologists and the absence of objective mechanisms to allocate psychologists to schools in Western Australia are briefly discussed. The next section provides background information on the construction of a measure of school need for psychological services. The validity of this exercise is argued by identifying examples of validity evidence using a seven component model proposed by Wolfe and Smith (2007a & 2007b). The findings are discussed in relation to the practical issue of providing psychological services to schools commensurate with an objective measure of school need.

Background

School psychology services in Western Australia

Written policy governing school psychology service provision is lacking in Western Australia. Similarly, there is no contemporary data about the functions of local practitioners and how they spend their time in schools. However, overseas school psychologists' services are well documented (Gilman & Gabriel, 2004; Oakland, Faulkner, & Annan, 2002). Although psychometric assessment of students' eligibility for education support placement and other specialist facilities is mandated in Western Australia by the Education Act (1999), and relevant legislative frameworks, the medical model orientation that is widespread in Western Australia perpetuates the problematic issues discussed by Sheridan & Gutkin (2000) and others (Oakland et al., 2002; Swerdlik & French, 2000). School psychologists elsewhere have shown a clear desire to reduce time spent on psycho-educational assessment (Gilman & Gabriel, 2004; Hosp & Reschly, 2002; Oakland et al., 2002; Reschly, 2000). In Western Australia, the preferred roles and functions of school psychologists have not been thoroughly studied or documented. In particular, the allocation of school psychologists to schools. This is despite the education reforms of the past decade including the Robson Report (2001) from the taskforce that investigated the structures, services and resources supporting Government schools. This report recommended that support services be closer to schools and allocated differentially to meet the diverse requirements of school leaders, teachers, students and their parents.

The empirical investigation of school need for psychological services

The aspects of school need for psychological services are posited to cluster around constructs extrapolated from the domains of service identified by Yesseldyke, Dawson, Lehr, Reschly, Reynolds, and Telzrow, (1997). Also in the delivery systems described by Oakland et al., (2002); the services delivered in Western Australia (Area Manager Student Services personal communication, 2009); and local Education District Student Services' Operational Plans. These aspects concern the characteristics of students, teachers and schools. They constituted the preliminary conceptual framework for the study and provided a foundation for the empirical investigation.

The instrument development process began with constructing a draft questionnaire based on the variables in the preliminary conceptual framework. A draft instrument was piloted with a randomly selected sample of four principals, three school psychologists and six teachers from two Department of Education districts in Western Australia. The results from this process were used to refine the instrument. Eventually, a 120-item pool of appropriate items was generated with multiple items written for each variable. The need to write so many items was because of the possibility of multi-dimensionality in the data which might have required constructing separate scales (the construct is comprised of three major sub-constructs); possibility of poor item-trait interaction leading to loss of items; a lack of surety about item difficulties and possible mis-targeting of items to persons. A four category Likert-type response scale was used (strongly agree, agree, disagree, and unable to judge). The items were tested in a small pilot study in which respondents commented on clarity of wording and ease of response.

Following revision, the instrument was trialled with 238 teachers from a random sample of twelve schools taken from the two education districts. To ease pressure on respondents, four versions of the instrument were administered with each containing a common set of 40 items and another 20 items unique to each form. The Rasch Unidimensional Measurement Model (RUMM) computer program (Andrich, Sheridan, Lyne & Luo, 2003) was used to test how well the data fitted the Rasch model, a requirement for measure construction. Then, a parsimonious scale of 35 items was developed by a stepwise process using item fit-statistics. The final set of items were chosen for good data-to-model fit, coverage of the construct domain and with a range of difficulties commensurate with person scores.

The resulting *Survey of Need for Psychological Services* (see Appendix) was a 35-item instrument using a four category response scale (strongly agree, agree, disagree and cannot judge). The 35 items were hierarchically arranged in seven sub-scales with five items in each of the sub-scales. The instrument was administered to principals and teachers in a stratified random sample of 18 schools from two Department of Education districts. Characteristics such as type of school and school size were represented as far as possible in the sample in the same proportion as they occur in the population of schools in the two districts. Data were analysed using RUMM2020 to provide summary test-of-fit statistics, category probability curves and threshold locations, individual item fit statistics, item characteristic curves and person-item distributions. These displays and statistics are estimates of how well obtained data fit the model. The results showed the data fitted the Rasch Rating Scale Model very well and enabled the researcher to plot the school scores successfully on the linear scale.

The validity of the measurement construction process was a key consideration. Provision of evidence to inform an argument for validity is the subject of this paper. The underlying theory is presented in the following section.

Validity Theory

The Kane (2001) review of validity theory identified three quite distinct conceptualisations of validity. The first, described in educational and psychological measurement literature in the early years of last century, emphasised the accuracy of estimations of the variable in question on the premise that greater accuracy indicated greater validity. This criterion-based view of validity was predicated on the assumptions that the criterion measure was equivalent to the value of the variable of interest and an appropriate criterion measure was readily available. This availability cannot be assumed. For example a psychological construct may not be directly observable or operationally definable. A solution to this problem was to engage an expert panel to make judgements about the content area covered by the test items with respect to the attribute of interest. This emerged as the second view of validity. The third approach was construct-based. This model in which measures were deemed to be observable exemplars of theoretical constructs, viewed validity as the extent to which measures behaved in a way that was consistent with extant theory (Wolfe & Smith, 2007a).

Thus three validity models were used. Each tended to be used for particular purposes without following any recognised principles to govern application. Ultimately, the introduction of the construct-based model led to Messick (1995) proposing three methodological principles for validation. These were: “an explicit statement of the proposed uses and interpretation of measures, extended analyses of multiple sources of evidence and the consideration of alternative theory-based interpretations of the meaning of measures” (Wolfe & Smith, 2007a, p. 98). The application of these principles to the other validity viewpoints resulted in a movement to develop a unified validity model, the construct-based approach provided the basis for unification.

Wolfe and Smith (2007b,) adopted Messick’s (1995) validation framework comprising six aspects of validity evidence in order to specify instrument development activities that would then lend support to subsequent validity arguments. They added the additional aspect of interpretability adopted from the Medical Outcomes Trust Scientific Advisory Committee (1995). The seven aspects of validity in their framework were:

1. Evidence of the content aspect;
2. Evidence of the substantive aspect;
3. Evidence of the structural aspect;
4. Evidence of the generalisability aspect;
5. Evidence of the external aspect;
6. Evidence of the consequential aspect; and

7. Evidence of the interpretability aspect. (p. 205)

These types of evidence were used to mount an argument for the validity of the measure construction process applied in the previous study.

Research Objectives
The aim of the study was to identify examples of evidence selected from the development process used to construct the linear scale to support arguments for the scale's validity. This included testing hypothesised relationships between dependent variable school need for psychological services and four independent school variables respectively. The research question was:

> To what extent do the RUMM2020 statistical analyses and graphical displays provide evidence to support an argument for the validity of the linear measurement scale?

Results
Scrutiny of the methodology and results of the measure construction process found six aspects of evidence. These are explained in the following section.

1. Evidence of the content aspect of validity
Evidence of the content aspect of validity includes explicit statements of the *purpose* of the study and the instrument development procedures. It also has to do with the relevance and ease of delineation of the construct upon which the items were to be based. The aim was stated explicitly in the formulation of three research questions, "*Can a linear scale be developed to measure school personnel perceptions of their school's need for psychological services? What aspects of the Wolfe and Smith (2007b) validity framework can be perceived in an investigation of school need for psychological services?* and *To what extent do the RUMM2020 statistical analyses and graphical displays constitute evidence of validity to support an argument for the validity of the linear interval scale?"* A paper based on this research was accepted by the Australian Association for Research in Education following a blind peer review process (see Young & Cavanagh, 2009).

Making the *domain of inference* explicit further supports the clarification of purpose. For example, the instrument development process was predicated on current school psychology theory and practice. The domain of inference was, in addition, criterion-based since it involved teacher perceptions, attitudes and school-specific behaviours. The type of inferences which may be drawn from the application of the instrument, which is one more condition of purpose, involved individual teachers and groups of teachers (that is, they were based in particular schools). A further example is the application of the final version of the scale in a number of targeted schools, of teacher and principal expression of school need for psychological services. Potential constraints such as the logistics of collecting data, time required to do so and availability of prospective respondents can inhibit the realisation of the purpose. However appropriate attention to the detail in the research design and the direct one to one relationships developed with school principals eliminated any potential constraints emerging during the study. These considerations were maintained throughout the instrument development process.

The *test specifications* focus on the procedures used in defining the construct model by describing the domain that is to be measured. The construct of *school need for psychological services* was deemed to be latent because it was not observable directly. A theoretical model was developed that specified the inter-related elements of the construct. This conceptual framework was critical for selecting the content and format of the items that would eventually constitute the substantive instrument. Test specifications also clarify the scoring model and the scaling model selected. A polytomous scoring model was applied and the Rasch Rating Scale Model (Andrich, 1988) provided the scaling.

Item development encompasses decisions to be made regarding the type of scale to be used, the number of response categories and how these response categories were to be labelled. In the case of the present study a three category response scale was used in all phases.

The content of the theoretical framework can be subject to expert review. The group also advised on clarity of wording of the individual items and the hierarchical arrangement of items within the sub-constructs. The experts were experienced teachers, school principals and managers of school psychologists. All were professionally well versed in school operations, student performance and school psychological services. Suggestions for improving all aspects of the theoretical framework and re-ordering of item hierarchies within the sub-constructs were accepted as evidence of Expert Review Group affirming the theoretical framework.

Evidence of the *technical quality of items* can be obtained by means of pilot studies and item trials. For example, the initial phase concerned item development and refinement processes. The subsequent phases involved survey item trials and instrument refinement. The quality of items can also be examined by means of statistical estimates showing the extent to which the distribution of observed values fits with values predicted by the measurement model. RUMM2020 estimates residuals for each item. A residual is the difference between an observed raw score and the score predicted by the Rasch Rating Scale Model.

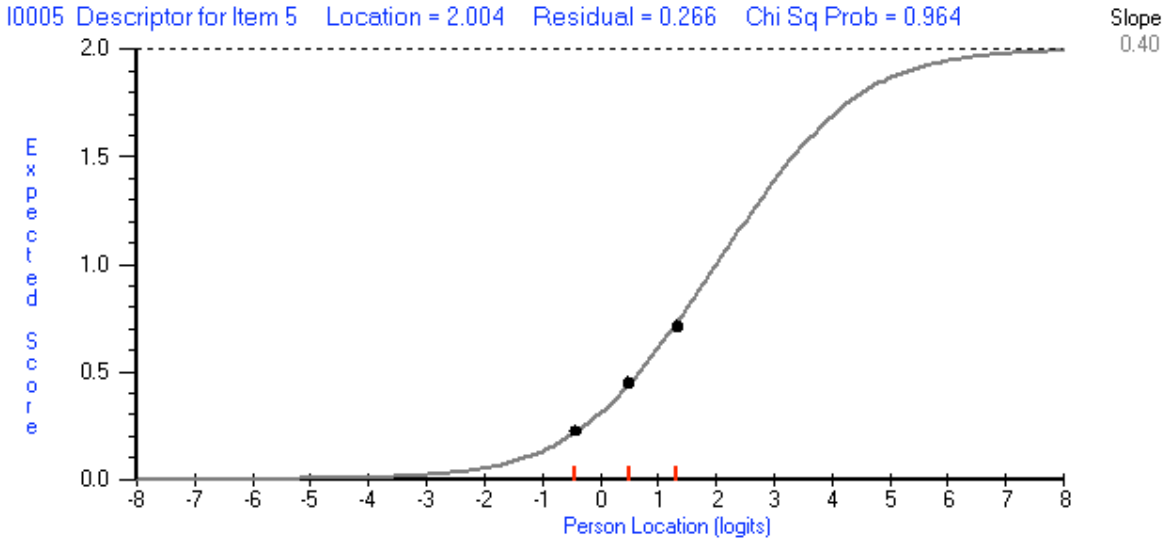

Figure 1: Item Characteristic Curve for Item 5 data

RUMM2020 also generates an Item Characteristic Curve (ICC) referred to as an ogive. The ogive illustrates the relationship between an item's expected value and person locations measured in logits. Observed scores for class intervals are then plotted along the ICC. For example, Figure 1 shows the ICC for Item 5 (*Test results are excellent)* eliciting data on school need for psychological services (N=148). The Rasch model curve increases smoothly as the location of persons increases with respect to the location of the item data as displayed in Figure 1. The logits on the horizontal axis are the logarithmetic odds of "need for psychological services" identified by school staff. The plots of the observed scores for three class intervals of school staff are appropriately close to the three values predicted by the model. The fit residual was 0.26 (<±2.5) because the observed scores were very close to the predicted scores. RUMM2020 sets a default value of ±2.5. Of the total 35 items 31 fit the model due to satisfying the specified criteria of having residuals <±2.50 and Chi Square probability values > 0.05. This indicates an excellent fit to the model and further

demonstrates the item technical quality contribution to the content aspect of validity evidence.

2. Evidence of the substantive aspect of validity

The substantive aspect of validity can be evidenced by the extent to which the theoretical framework, or *a priori* rationale, can explain any observed consistencies among item responses. An example is confirming hypothesised item difficulty hierarchies within the sub-constructs that comprise the instrument. With respect to the present study the theoretical model informing the study suggested that school need for psychological services not only subsumed a number of sub-constructs but that the items within them had a hierarchical arrangement. Although the conceptual model indicated a particular sequence of items in ascending order of difficulty the order was quite subjective rather than empirically determined. Item fit statistics illustrating sub-scale item sequence and item difficulty are given in Table 1 below.

Table 1
Item sequence and difficulty

Item	Sub-construct item sequence in the model	Location	SE
	Teaching		
1	Teachers cater for individual differences	-1.01	0.16
2	Student progress is documented regularly	-0.68	0.17
3	Teachers know what each student needs	-0.00	0.15
4	Teaching and learning produces high achievement	0.88	0.15
5	Test results are excellent	2.00	0.16
	Development of academic skills		
6	Students need extra help	-1.78	0.17
7	Students like to learn	0.06	0.16
8	Students respond well	0.24	0.18
9	Students are attentive	1.74	0.17
10	Students access study skills training	3.97	0.24
	School development of socialization/life skills		
11	Student attitudes are important	-1.76	0.18
12	The school rewards appropriate behaviour	-1.46	0.16
13	Behavioural issues are well managed	-0.97	0.16
14	There are few discipline problems	1.09	0.13
15	Students quickly resolve conflict	1.88	0.15
	Inclusivness in learning and development		
16	All classes have students with learning difficulties	-1.23	0.16
17	We welcome students from diverse backgrounds	-1.68	0.17
18	Teachers celebrate the school's diversity	-0.80	0.15
19	New students can be seen by the psychologist	0.77	0.15
20	We use the psychologist's ideas for our programs	0.76	0.15
	Prevention services and wellness promotion		
21	There is a need for child protection training	-0.65	0.14
22	Programs have improved student well-being	-0.81	0.20
23	The school has suicide prevention strategies	2.63	0.24
24	Parents utilise healthy eating programs	1.58	0.16
25	The school coordinates mental health services	1.08	0.19

	Home/school/community collaboration		
26	Parents are welcomed into the school	-2.86	0.18
27	The school keeps the community informed	-1.71	0.17
28	The community helped develop the school ethos	-0.10	0.17
29	Parents are active in the School Council	0.27	0.13
30	Teachers find parents easy to engage	1.07	0.13
	School Climate		
31	Teachers strive for school improvement	-1.64	0.17
32	Teachers provide agenda items for staff meetings	-0.45	0.15
33	Leadership is shared among teachers	-0.26	0.14
34	Teacher/psychological consultation is in place	0.06	0.15
35	Psychological services improve school climate	-0.22	0.16

Scrutiny of the RUMM2020 estimates of item difficulty revealed that the order of measured item difficulty showed a close correspondence with the order of items as they were originally presented in the conceptual model. In other words, the order of difficulty of items in general is as initially hypothesised. This relationship therefore constitutes evidence for the substantive aspect of validity.

RUMM2020 produced a Category Probability Curve for each of the items. Figure 2 below shows the Category probability curve for Item 3 data (*Teachers know what each student needs*).

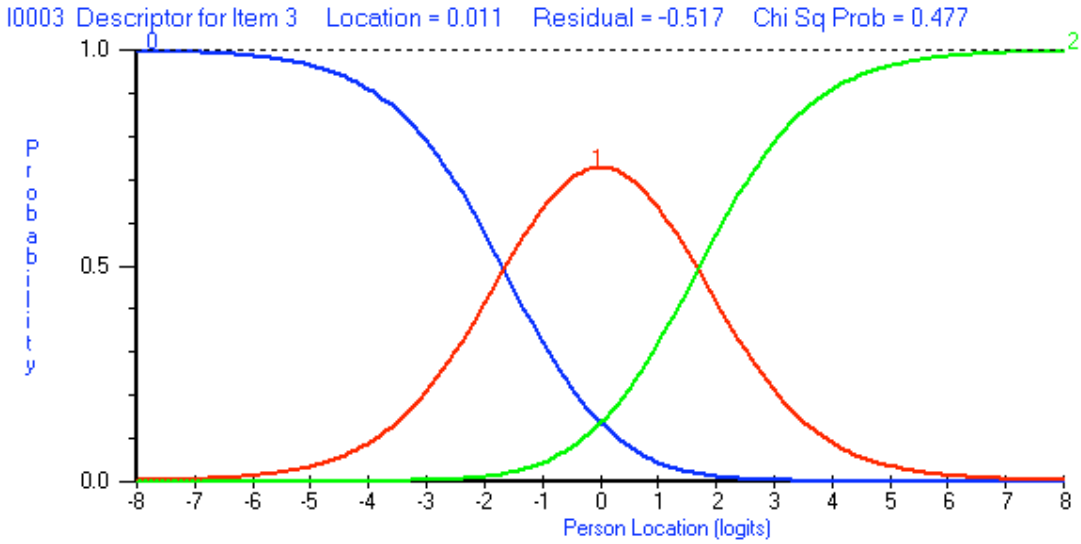

Figure 2: Category probability curve for Item 3 data

The probability of a category being chosen is plotted on the vertical axis and person location (school need for psychological services measured in logits) plotted on the horizontal axis. Persons with ability to affirm more difficult items are located to the right of the horizontal axis and those with less ability, lie to the left. Curve 0 (disagree) shows that the probability of a person located 8 logits below the mean selecting category 0 is 1.0. The probability of selecting this category decreases to zero for persons located +2.3 logits above the mean. For curve 1 (agree) the probability increases from 0.00 for a person located 6 logits below the mean to a maximum of about 0.75 for persons located at the mid-point of the person location scale. The probability then decreases to zero for persons located + 6.0 logits above the mean.

Item thresholds were estimated to identify the person ability measure at which an equal probability existed for persons to select either one of two contiguous response categories.

The intersection of Curve 0 and Curve 1 is the threshold for the disagree and agree categories with a value of -1.66 logits. The other threshold (agree and strongly agree) for Curve 1 and Curve 2 is +1.68 logits. A distinctive feature of the model is that "it can reveal if the ordering of the categories is not working as intended" (Andrich & Styles, 2004, p. 9). For any particular item the probability of choosing response categories ought to be ordered in line with the person ability values expressed in logits. Figure 2 shows that the thresholds are properly ordered. All 35 items were characterised by ordered thresholds indicative of response categories being selected in a logical and consistent manner by school personnel. The agreement between theoretically-based expectations and observed item functioning is additional evidence for the substantive aspect of validity.

3. Evidence of the structural aspect of validity

The structural aspect of validity is concerned with the construct domain and the adopted scoring model. For example, by confirming if the demands of a unidimensional measurement model are realised when a unidimensional trait is measured. As far as unidimensionality is concerned a RUMM2020 analysis of residuals achieves this by first extracting the linear measure component from the data set assuming that all items are equally weighted on the first component. Then a Principal Components Factor analysis is performed on the standardised residuals after the initial Rasch scaling. This process was applied to the 35 item data. Table 2 shows the percentage of the total variance accounted for by each of the principal components (eigenvalues). For example, for the first item 4.63/35=0.1324 or 13.24%. One component accounted for 13.24% of the total variance which indicates the possibility of multidimensionality within the data.

Table 2
Principal Components Summary

Principal Component	Eigen value	Percentage	Cumulative percent	Standard error
PC001	4.63	13.24%	13.24%	0.65
PC002	2.73	7.82%	21.06%	0.38
PC003	2.48	7.11%	28.17%	0.34
PC004	2.01	5.75%	33.92%	0.27
PC005	1.95	5.58%	39.50%	0.27
PC006	1.68	4.82%	44.32%	0.23
PC007	1.50	4.30%	48.62%	0.20
PC008	1.36	3.90%	52.52%	0.18
PC009	1.29	3.70%	56.21%	0.17
PC010	1.22	3.50%	59.71%	0.17
PC011	1.19	3.40%	63.11%	0.16
PC012	1.06	3.04%	66.15%	0.14
PC013	1.01	2.91%	69.06%	0.13
PC014	0.93	2.68%	71.74%	0.13
PC015	0.89	2.55%	74.28%	0.12
PC016	0.85	2.44%	76.72%	0.11
PC017	0.78	2.24%	78.96%	0.10
PC018	0.76	2.19%	81.15%	0.10
PC019	0.70	2.01%	83.16%	0.09
PC020	0.67	1.94%	85.10%	0.09
PC021	0.65	1.86%	86.96%	0.09
PC022	0.60	1.72%	88.68%	0.08
PC023	0.54	1.56%	90.23%	0.08
PC024	0.49	1.41%	91.64%	0.07
PC025	0.47	1.36%	93.00%	0.07
PC026	0.42	1.22%	94.22%	0.06
PC027	0.39	1.13%	95.35%	0.06
PC028	0.33	0.96%	96.31%	0.05

PC029	0.31	0.90%	97.21%	0.05
PC030	0.28	0.82%	98.03%	0.05
PC031	0.25	0.73%	98.76%	0.05
PC032	0.24	0.71%	99.47%	0.05
PC033	0.19	0.56%	100.03%	0.04
PC034	0.10	0.30%	100.33%	0.05
PC035	-0.11	-0.33%	100.00%	0.03

4. Evidence of the external aspect of validity

The external component of validity may be the most important (Wolf & Smith, 2007a) as it examines the extent to which the interval scale measure (dependent variable) is related to external or independent measures of the same or similar constructs. In the case of the present research the external measures were *student suspensions, student absenteeism, school socio-economic index and students with individual behaviour plans*.

Publicly available school-level data were collected such as socio-economic index, student suspension and exclusion data, truancy data and numbers of students with individual behaviour management plans. Quotients were calculated for absenteeism, suspensions and students with individual behaviour plans by dividing the raw data obtained from the Department of Education by the student population of each school. This effectively removed the extraneous variable of school student population size from the analyses. These were treated as independent variables and the linear measure of school need for psychological services was treated as the dependent variable.

Initially, correlations between variables were estimated. Then, a multiple regression analysis with school need for psychological services and school demographic data being treated as the dependent and independent variables respectively was then carried out. Hypothesised relationships between the linear measure and the four school-level variables were then investigated by multiple regression. Confirmation of a relationship needed a probability level of less than 0.01 that the relationship was attributed to random variations in the data. The direction of the relationship (direct or inverse) was indicated by the positive or negative value of the slope coefficient (b). The effect of variation in the independent variables on variation in the dependent variable (the strength of association) was computed for each relationship. The beta weight (β) was calculated to provide a standardised measure of the strength of association between each of the four independent variables and the dependent variable. The cumulative effect of variation in the independent variables on the dependent variable was measured by computing R^2.

The first result from the correlational analyses is the presence of a number of inverse and positive associations between the five variables (see Tables 3 and 4).

Table 3
Correlation between independent variables

	Correlation coefficient			
Independent variable	1	2	3	4
1.School socio-economic index decile	1.00			
2.School absenteeism quotient	−0.55*	1.00		
3.School suspensions quotient	−0.29*	0.10	1.00	
4.Students with individual behaviour plans	−0.41*	0.00	+0.25*	1.00

Note: *p < 0.01

These statistically significant inverse correlations show that as *school socio-economic index* increases, *school absenteeism, school suspensions* and *number of students with individual behaviour plans* decrease. The positive correlation (+0.25) indicates that as *school suspensions* increase, the *number of students with individual behaviour plans* also increases.

Table 4
Correlation between independent variables and the dependent variable

Independent variable	Dependent variable Measure of school need for psychological services
1.School socio-economic index decile	0.02
2.School absenteeism quotient	−0.09
3.School suspensions quotient	−0.30*
4.Students with individual behaviour plans	−0.08

Note: *p < 0.01

Three of the four independent variables showed relations with the dependent variable that were not statistically significant. However, there was a statistically significant inverse relation between *school suspensions* and the measure of *school need for psychological services*. This relation indicates that as *school suspensions* increase the *school need for psychological services* also increases.

It is also necessary to examine associations between the dependent variable and multiple independent variables as a group, since this could produce different effects. Multiple regression analysis was applied to test for a relationship between the dependent variable and the four independent variables. The independent variables were step-wise regressed against the dependent variable. The four independent variables accounted for 11% (R^2 = 0.11) of the variance in *School need for Psychological Services* (dependent variable) as shown in Table 5.

Table 5
Regression analysis

Independent variables	b	Std error of b	β	t	p
1.School socio-economic index decile	0.03	0.03	−0.13	−1.25	0.212
2.School absenteeism quotient	−0.02	0.01	−0.14	−1.41	0.160
3.School suspensions quotient	−1.39	0.34	−0.33	−4.04	0.00*
4.Students with individual behaviour plans	−0.14	0.70	−0.19	0.21	0.832
Constant	1.02	0.27		3.77	0.000

R = 0.33
R^2 = 0.11
F = 4.52 p < 0.002
Note*p < 0.01

School need for Psychological Services was confirmed to have s strong inverse relationship with *school suspensions* (β = - 0.33, p < 0.01). That is, when there is a unit positive change in *school suspensions* then there is a decrease of 0.33 in the measure of *school need for psychological services*. In simple terms, an increase in student suspensions leads to an increase in the need for psychological services when the other independent variables are mutually controlled. The small inverse relationships between the dependent variable and the other three independent variables (*school socio-economic index decile; school absenteeism quotient; students with individual behaviour plans*) were not statistically significant. These

results show one strong independent association between the *school need for psychological services* measure and the school demographic variable of *school suspensions*. This finding provides partial support for the hypothesis concerning associations between the school demographic data and the measure of *school need for psychological services.*

5. Evidence of the consequential aspect of validity

The consequential aspect of validity focuses on considerations about the ways in which interpretations of scores may be of consequence. For example, the consequences of specifying individual school logit scores on the linear scale to principals. Reporting to client principals used tables, graphical displays and textual data to qualify the characteristics of their school and staff scoring in the bottom 20% of the sample. This categorisation was used to exemplify the attributes of schools with high need for psychological services.
A further consequence was the risk of this data becoming publicly available particularly if detrimental to any school. Steps were therefore taken to avoid this by allocating a code to each school and maintaining the anonymity of participating staff.

6.Evidence of the interpretability aspect of validity

The interpretability aspect of validity is defined as “the degree to which qualitative meaning can be assigned to quantitative measures” (Wolfe & Smith 2007a, p. 100). This has of course implications for the quality of the communicative method chosen. One method of interpreting Rasch generated data is the use of graphical displays such as the Person-Item Threshold distribution shown in Figure 3.

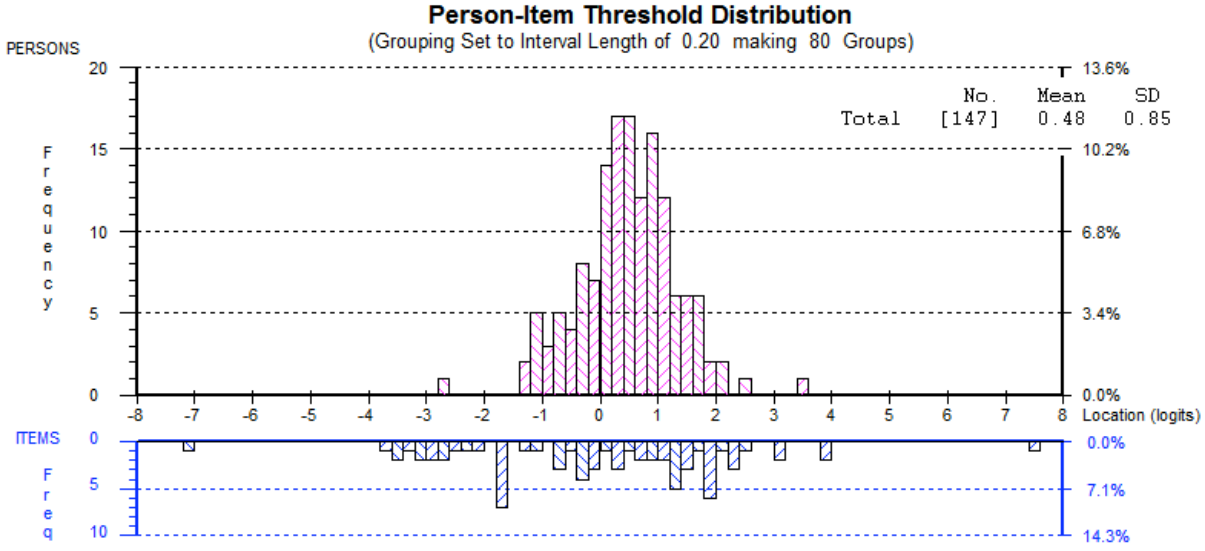

Figure 3: Person-Item threshold distribution for 35-item scale

The display represents degree of difficulty an item presented to teachers on the bottom scale and teacher ability to affirm increasingly more difficult items on the top scale. The logarithmic odds scale shows both item difficulty values from -7.2 logits (easy to affirm) to + 7.4 logits (more difficult to affirm) and teacher ability measures (from -2.8 logits (lower ability) to +3.4 logits, (higher ability). The display facilitates easy comparison of individual respondent scores with the difficulties of items. In addition it demonstrates the extent to which the item difficulties targeted or matched the abilities of the respondents. There were no obvious gaps in the range of item difficulty within the teacher ability continuum. The relationship between these two metrics demonstrates that the items and person abilities are well matched and therefore the persons are well targeted by the scale items with a satisfactory number of easy and hard items across the scale. The ease of reading the data presented above is evidence of the interpretability aspect of validity.

Conclusion

Evidence from the development activities and results of a quantitative investigation were used to demonstrate six aspects of validity. The utilisation of the Rasch Rating Scale Model clearly generated comprehensive evidence for the validity of the *Survey of Need for Psychological Services*. In addition, a significant inverse association between the school measures and school suspension rates was observed. As the research intentions were realised, informed decisions can now be made about the level of psychological services that should be provided to schools, congruent with the psychological needs of their students.

References

Andrich, D. (1988). *Rasch Models for Measurement.* Series: Quantitative Applications in the Social Sciences, Series number 07-068. Sage Publications Inc. London, Delhi.

Andrich, D., Sheridan, B., Lyne, A. & Luo, G. (2003). RUMM: *A windows-based item analysis program employing Rasch unidimensional measurement models*. Perth: Murdoch University.

Andrich, D. & Styles, I. (2004). *Final report on the psychometric analysis of the Early Development Instrument (EDI) using the Rasch model: A technical paper commissioned for the development of the Australian Early Development Instrument (AEDI).* Perth: Murdoch University.

Gilman, R. & Gabriel, S. (2004). Perceptions of School Psychological Services by Education Professionals: Results from a Multi-State Survey Pilot Study. *School Psychology Review, 33* (2), 271-286.

Hosp, J. L., & Reschly, D. J. (2002). Regional differences in school psychology practice. *School Psychology Review, 31* (1), 11-24.

Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement, 38*(4), 319-342.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' Responses and performances as scientific enquiry into score meaning. *American Psychologist*, 50, 741-749.

Oakland, T., Faukner, M., & Annan, J. (2002). School Psychology in Four English-speaking Countries: Australia, Canada, New Zealand and the United States. In C.Frisby and C. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology*. New York: Wiley & Sons.

Reschly, D. J. (2000). The present and future status of school psychology in the United States. *School Psychology Review, 29*(4), 507-522.

Robson, Alan. (2001) *Investing in Government schools: Putting children first. Report of the taskforce on structures, services and resources supporting Government schools.* Perth: Department of Education and Training.

Government of Western Australia (1999). School Education Act. Perth:author.

Sheridan, S. M., & Gutkin, T. B. (2000). The ecology of school psychology: Examining and changing our paradigm for the 21st Century. *School Psychology Review, 29*(4), 485-502.

Swan Education District Student Services Operational Plan 2009-2011. Perth: Department of Education and Training.

Swerdlik, M. E., & French, J. L. (2000). School psychology training for the 21st Century:

Challenges and opportunities. *School Psychology Review, 29*(4), 577-587.
West Coast Education District Student Services Operational Plan. (2009-2010). Perth: Department of Education and Training.
Wolf, E. W. & Smith, E. V. (2007a). Instrument Development Tools and Activities for MeasureValidation using Rasch Models: Part 1-Instrument Development Tools*. Journal of Applied Measurement, 8*(1), 97-123.
Wolf, E. W. & Smith, E. V. (2007b). Instrument Development Tools and Activities for Measure Validation using Rasch Models: Part 2-Validation Activities*. Journal of Applied Measurement, 8*(1), 204-234.
Yesseldyke, J., Dawson, P., Lehr, C., Reschly, D. J., Reynolds, M., & Telzrow, C. (1997). *School Psychology: A blueprint for training and practice* II. Bethesda, MD: National Association of School Psychologists.
Young, A. & Cavanagh, R. F. (2009). *A Rasch Modelling Approach to Measuring School Need for Psychological Services.* Paper presented at the 2009 Annual Conference of the Australian Association for Research in Education: Canberra.

Appendix

SURVEY OF NEED FOR PSYCHOLOGICAL SERVICES

Your school is:

INSTRUCTIONS:

Please do **NOT** write your name just remember the number in the top right corner.

If you **strongly agree** with the statement, please circle **4**

If you **agree** with the statement, circle **3**

If you **disagree** with the statement, circle **2**

If you **can't judge**, circle **1**

Teaching	Strongly agree	Agree	Disagree	Can't Judge
1.Teachers cater for individual differences	4	3	2	1
2.Student progress is documented regularly	4	3	2	1
3.Teachers know what each student needs	4	3	2	1
4.Teaching and learning produces high achievement	4	3	2	1
5.Test results are excellent	4	3	2	1

Development of academic skills	Strongly agree	Agree	Disagree	Can't Judge
6.Students need extra help	4	3	2	1
7.Students like to learn	4	3	2	1
8.Students respond well	4	3	2	1
9.Students are attentive	4	3	2	1
10.Students access study skills training	4	3	2	1

School development of socialization and life skills	Strongly	Agree	Disagree	Can't Judge

	agree			
11.Student attitudes are important	4	3	2	1
12.The school rewards appropriate behaviour	4	3	2	1
13.Behavioural issues are well managed	4	3	2	1
14.There are few discipline problems	4	3	2	1
15.Students quickly resolve conflict	4	3	2	1

Inclusiveness in learning and development	Strongly agree	Agree	Disagree	Can't Judge
16.All classes have students with learning difficulties	4	3	2	1
17.We welcome students from diverse backgrounds	4	3	2	1
18.Teachers celebrate the school's diversity	4	3	2	1
19.New students can be seen by the psychologist	4	3	2	1
20.We use the psychologist's ideas for our programs	4	3	2	1

Prevention services and wellness promotion	Strongly agree	Agree	Disagree	Can't Judge
21.There is a need for child protection training	4	3	2	1
22.Programs have improved student well-being	4	3	2	1
23.The school has suicide prevention strategies	4	3	2	1
24.Parents utilise healthy eating programs	4	3	2	1
25.The school coordinates mental health services	4	3	2	1

Home/school/community collaboration	Strongly agree	Agree	Disagree	Can't Judge
26.Parents are welcomed into the school	4	3	2	1
27.The school keeps the community informed	4	3	2	1
28.The community helped develop the school ethos	4	3	2	1
29.Parents are active in the School Council	4	3	2	1
30.Teachers find parents easy to engage	4	3	2	1

School climate	Strongly agree	Agree	Disagree	Can't Judge
31.Teachers strive for school improvement	4	3	2	1

32.Teachers provide agenda items for staff meetings	4	3	2	1
33.Leadership is shared among teachers	4	3	2	1
34.Teacher/psychologist consultation is in place	4	3	2	1
35.Psychological services improve school climate	4	3	2	1

THANK YOU VERY MUCH FOR TAKING THE TIME TO COMPLETE THIS SURVEY