Confirmatory Factor Analysis of the Westerman Aboriginal Symptom Checklist – Youth (WASC-Y)

by Jonathon Little (2007)

Acknowledgement is given to Dr. Peter Sevastos (Curtin University of Technology) and Dr. Craig Enders (Arizona University) for their advice on the statistical analysis.

The WASC-Y has been developed as a structured 53-item (originally 56-item) scale designed to obtain self-report perceptions from 13- to 17-Year-old Aboriginal youth regarding their typical behaviours and feelings. The scale was designed as a paper and pencil individual or group administration, but is also able to be administered orally to youth that have low level literacy skills. For an overview of the original validation of the WASC-Y scale please refer to the original WASC-Y manual (see Westerman, 2003).

Items on the WASC-Y are grouped into seven separate subscales and purport to represent the following mental ill health and constructs in Aboriginal youth;

1. Depression (11 items)

2. Suicidal behaviours (10 items)

3. Alcohol and drug usage (2 items)

4. Impulsivity, hyperactivity and agitation (3 items)

5. Anxiety (11 items)

6. Self-esteem (8 items)

7. Resilience to Mental Ill Health (11 items)

The WASC-Y presents as a series of likert scales. Questions 1 – 13 and 19 – 56 are five-point scales (1 = "Never"; 2 = "Little Bit"; 3 = "Half and Half"; 4 = "Fair Bit", and 5 = "Heaps"). Items 15 and 18 are a "True" (1) or "False" (2) option. Respondents are required to mark those items on the inventory that they judge to be most descriptive of the way they "usually" feel and think about life. Please refer to Chapter Twelve for specific instructions provided to youth during administration of the WASC-Y.

This Chapter describes the statistical analysis that has been involved in the updated validation study and provides contextual information from the initial validation study of Westerman (2003). This section will result in (1) finalisation of the WASC-Y measurement model, and (2) determining the internal consistency of the WASC-Y subscales, (3) determining whether the WASC-Y represents the constructs (of depression, suicide, hopelessness, impulsivity, anxiety, self-esteem and protective factors) that have been hypothesised. This has been determined through Factor Analysis (FA).

Introduction to factor analysis

Factor analysis (FA) is based on the assumptions of the common factor model (Lord & Novik, 1968) where indicators (the terms indicators, items or observables are used interchangeably) are dependent on an unobservable latent variable or variables (commonly referred to as a latent construct or factor). Latent constructs are not directly observable; they can only be measured indirectly with some degree of error. Implied under the assumptions of the common factor model is that indicators represent the effects of a latent construct, that is, they are caused by it and hopefully nothing else, and indicators of this type have been referred to as effect indicators or more broadly they belong to effect indicator measurement models (Bollen & Lennox, 1991). Factor analysis offers insight into how many latent constructs explain the covariation among a set of observables. At times the factors may represent constructs that are of theoretical interest such as depression or intelligence, while at other times, these factors may represent systematic sources of bias, such as social desirability or other response sets. There are two basic types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). For an extensive discussion of EFA the reader is referred to the excellent review of Fabrigar, Wegener, MacCallum and Strahan (1999) and also Thompson (2004). For an extensive discussion of CFA see Bollen (1989) and for an introductory text, see Raykov and Marcoulides (2000) and Schaumaker and Lomax (2004).

EFA is used when the relationships between the observables and their underlying latent factors are uncertain. No prior assumptions are made about which items reflect any specific latent construct and each item is free to load on any factor. Further, no assumptions are made about the number of factors that may be present. The advantage of using exploratory factor analysis, as opposed to confirmatory factor analysis, is that it enables detection of any unhypothesised cross-loadings. That is, items which inaccurately reflect the hypothesised construct that is specified. EFA is a model generating process and consideration is often given to the number of factors that may be present. True exploratory analyses are rare in that researchers will often have some *à priori* assumptions about what latent constructs are of influence. The use of maximum likelihood exploratory factor analysis (ML EFA) represents a progression beyond pure EFA in that a hypothesised number of factors can be extracted and the goodness-of-fit of this model to the sample data can be tested.

CFA represents the next progression from model generating to model testing. Unlike EFA, CFA is used when the researcher has some idea of the underlying latent structure. Items are specified to load onto their respective factors *à priori* while other parameters would be restricted to values on non-target factors that are typically zero (although zero loadings are somewhat unrealistic). CFA then, is theory driven in contrast to EFA which is data driven. Ideally, a sample would be randomly split and an EFA conducted on one sub-sample. This EFA would then be used to generate a measurement model that would then be cross-validated on an independent hold out sample using confirmatory factor analysis. To be strictly confirmatory the parameter values would have to be constrained to the values previously estimated. The goodness-of-fit of this model to the sample data would then be tested statistically. In many research contexts, data availability precludes sample splitting and an EFA and CFA are sometimes conducted on the same sample data. While the current research will use a confirmatory factor analysis to test the fit of the proposed measurement model outlined by Westerman (2002) this approach will not be strictly confirmatory. Rather, the sample used will consist of approximately 334 cases of which 189 (56%) were used in Westerman's original analysis. Analyses of this type are not strictly confirmatory and are conditional until cross-validated on an independent sample.

The proceeding sections will outline the rationale behind the statistical decisions made in the analysis of the WASC-Y. Following this, results to a CFA of the WASC-Y subscales will be reported and a discussion will highlight potential paths for future development and inquiry.

Method

This section provides thorough details regarding the rationale behind analytical decisions and the application of the various statistical techniques used in the confirmatory factor analysis. This includes a discussion of the assessment of model fit, statistical power, multivariate normality and outliers, the management and analysis of missing data, reliability, convergent and discriminant validity, and model modification.

Assessment of model fit

The most common indicator of model fit is the chi-square statistic. Ordinarily a small chi-square to degree of freedom ratio is sought, typically less than 2 or 3 (Kline, 1998). It is widely acknowledged that the likelihood ratio test is sensitive to sample size, and a huge variety of fit indices have been developed, in part to address the χ^2 limitations. Following Byrne (2006) and the recommendations of Hu and Bentler (1995) three measures of fit are used in the current research. These include the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Standardised Root Mean Square Residual (SRMR).

The RMSEA "…estimates the lack of fit in the model compared to a perfect (saturated) model" (Ullman, 2000, p. 699). As the goodness-of-fit improves the RMSEA and SRMR approach their lower bound value and reach zero when the model fits perfectly (Brown, MacCallum, Kim, Anderson & Glaser, 2002). Confidence intervals for the RMSEA can be produced and used to estimate the precision of the discrepancy value. Point estimates of the RMSEA less than .05 constitute good fit, values between .05 and .08 correspond to an adequate fit, and values between .08 and .10 represent mediocre fit, while values in excess of .10 represent poor fit (Byrne, 2006). An assessment of the RMSEA statistic should be considered in the context of the statistic's confidence interval and this is discussed in greater detail below with regard to the RMSEA closeness of fit test.

The SRMR "…represents the average discrepancy between the observed sample and hypothesised correlation matrices [meaning that the model] …explains the correlations to within an average error" (Byrne, 2006, p.99) corresponding to the SRMR statistic. SRMR values less than .08 are desirable (Hu & Bentler, 1999). The CFI compares the fit of the model with the null model which assumes that the latent variables are uncorrelated. While original recommendations suggested that CFI values > .90 usually represent well-fitting models (Bentler, 1992) CFI values > .95 are now recommended (Hu & Bentler, 1999).

In the context of maximum likelihood EFA the Expected Cross Validation Index (ECVI; Browne & Cudeck, 1989) is used to compare nonnested models. There exist no guidelines regarding the interpretation of ECVI's magnitude, although lower values are desirable.

Power of the RMSEA closeness of fit test

The assessment of model fit requires that there is sufficient power "for detecting when a hypothesis about model fit is false" (MacCallum, Browne & Sugawara, 1996, p. 138). Power then, can be defined as "…the likelihood of drawing the correct conclusion about a false null hypothesis regarding model fit". In testing the hypothesis of close fit, the power is calculated in those instances where the RMSEA closeness of fit is rejected. It is important to recognise that while it may be the case that the sample size may be more than sufficient to ensure an acceptable level of power for the RMSEA closeness of model fit test, it does not necessarily follow that this sample size will be large enough to "…obtain parameter estimates that have standard errors small enough to be of practical value" (Anderson & Gerbing, 1988, p. 415).

Multivariate normality and outliers

Maximum likelihood estimates of model parameters and tests of goodness of fit assume that the manifest variables have a multivariate normal distribution (West, Finch & Curran, 1995). Prior to fitting any model, it is prudent to investigate deviations from multivariate normality and search for univariate and multivariate outliers. This is important because "even if a proposed structure is correct for the majority of the data in a sample, a small proportion of outliers leads to biased estimators and significant test statistics" (Yuan & Bentler, 2001, p. 161). Mardia's (1970) test for multivariate skewness and of particular interest, multivariate kurtosis is used in the current research together with Yuan, Lambert and Fouladi's (2004) estimate of multivariate kurtosis for use with missing data. Univariate outliers are identified as z scores larger than 3.64 at $p < .001$, two tailed (Tabachnick & Fidell, 2001). Multivariate outliers are identified using Mahalanobis's distance at $p < .001$.

Satorra and Bentler (1988) developed a scaling correction to the χ^2 statistic (S-B χ^2) for use with data that was not multivariate non-normal. When using this robust approach, standard errors and test statices are corrected and robust versions of the CFI and RMSEA (including its 90% confidence interval) are produced, hereafter these are identified as *CFI and *RMSEA. When data are both incomplete and non-normal the Yuan and Bentler (2000) scaled statistic χ^2 statistic (Y-B χ^2) is used. Both the S-B $\chi^2$and the Y-B χ^2 are used in the current analysis.

Two final points are worth noting in relation to multivariate normality and outliers. First, in the current research factor analyses are replicated, both with outliers retained and with outliers either recoded and/or removed. As Thompson (2004) argues: "If interpretations are robust across various decisions, at least the researcher can be confident that the results are not artefacts of methodological choices." (p. 125).

During estimation the management of outliers should not require exhaustive scrutiny as estimation constitutes only one of various stages before the final estimation of model parameters and goodness of fit (Cohen, Cohen, West & Aiken, 2003).

Asymptotically distribution free (ADF) estimation methods have been claimed by some (e.g., Browne, 1984) to be insensitive to the distributions of the observed variables, however these techniques require extremely large sample sizes (West, Finch & Curran, 1994) – easily in excess of 2,500. Numerous studies have found that the ADF estimation method produces χ^2 statistics that are far too high when sample sizes are small to moderate. Muthén (1984) developed an alternative method CVM (continuous/categorical variable methodology) which can be applied to ordered categories through the use of polychoric correlation when correlated variables are of an ordinal scale. While the CVM method is able to provide appropriate estimates of the model χ^2, parameter estimates and their standard errors, West *et al.,* (1995) suggested that there is little or no benefit in using this technique with items that have 5 or more anchors (such as those in the WASC-Y). More recent findings have also supported this assertion and show that the χ^2 is most severely affected when two response categories are used (Green, Akey, Flemming, Herhberger & Marquis, 1997).

Missing Values

Making valid and efficient inferences of population parameters pose challenges when not all data elements are observed. The task when encountering missing values is "…not to estimate, predict, or recover missing observations nor to obtain the same results that we would have seen with complete data" (Schafer & Graham, 2002, p. 149). Rather, the purpose is to obtain valid and precise estimates of population parameters. According to Little and Rubin (1989), there are two patterns of missing values, which can be distinguished: 1) missing at random (MAR), and 2) missing completely at random (MCAR). Missing at random means that, the probability that Y is missing is contingent on observed values of X but not contingent on unobserved values of Y.

A more stringent form of MAR, MCAR, assumes that the probability of Y being missing *cannot* be contingent on observed values of X nor unobserved values of Y. If it is assumed that MAR holds, independent samples *t*-tests can be used to assess the MCAR assumption. Variables with missing data (usually those with more than 5% missing observations) would be split into two nominal groups (missing vs. nonmissing). The null hypothesis is that no differences would be observed for any item across these two groups (missing vs. nonmissing). Note that the *t* tests will not be independent and therefore their probability levels are questionable. More sophisticated tests of the MCAR assumption can also be assessed using Little's homogeneity of means test in which the null hypothesis is that data is MCAR (Little, Roderick & Schenker, 1995) and Kim and Bentler's (2002) generalised least squares (GLS) tests for the homogeneity of covariance matrices and a combined test for the homogeneity of means and covariances matrices is also available.

The analysis of the suicide subscale was complicated by the use of a skip rule for item 12. There were 5 items (items 13 to 17) which respondents were only required to score if they provided a positive response to item 12. For instance question 12 "I have felt so sad I have thought of ending my life" – a response of "Never" *may* mean that questions 13 – 17 are no longer applicable as they assume some form of suicidal ideation.

An analysis of the validity of the skip rule for item 12 found that 30 (20.87%) respondents to item 13 also answered "never" (1) on item 12, rather than skipping items 13 to 17 and proceeding to item 18 as directed. Further examination of cross tabulations of item 12 with items 14 to 17 revealed a similar pattern of responding. Further complicating the analysis of the suicide subscale is that missing responses to the item 13 through 17 are causally contingent on responses to item 12. That is, the missing data on items 13 to 17 were not missing completely at random (indeed they are partly missing by design) as defined in Rubin's terms (1987). They are by definition MAR. Statedly differently, the patterns of missing data were designed to provide different estimates of the population means and the population covariance matrixes as item 12 classifies (within an uncertain degree of reliability) two populations of individuals, those with no suicidal thoughts versus those with some amount of suicidal thoughts.

Kim and Bentler's (2002) generalised least squares test of homogeneity of means and covariance matrices was used to test this assumption and as expected, this test was statistically significant, χ^2 (208, N = 326) = 509.208, p = 2.874E-27. From this test, the MCAR hypothesis that the various patterns of missing data do not provide different estimates of the population means and the population covariance matrices is rejected.

An analytical strategy for handling the missing data for the suicide subscale would be to use multiple imputations to estimate missing values on items 12 through 17 under the MAR assumption. In the current research, the use of multiple imputations was considered computationally burdensome and unrealistic; in addition there are no rules for combining fit indices across imputations (Enders, in press). Two alternatives to the analysis of the suicide subscale include the use of the expectation maximisation (EM) algorithm type of maximum likelihood (Yuan & Bentler, 2000), or under the assumption that missing data is missing by design, a multi-sample CFA (Bentler, 2006).

If the EM algorithm is used by including item 12 as either a manifest indicator or as an auxiliary variable the MAR assumption for items within the suicide subscale should hold. Alternatively, the EQS multi-sample technique could be used to treat each distinct pattern of missing observations as a subsample with equality constraints on parameters across each subsample. There are 13 patterns of missing data observed across items 12 through 21 (excluding items 15 and 18). All but two patterns constitute less than 2% of all observations, although 170 cases (51.20%) have no missing values while one pattern consists of 139 cases (41.87%) on which missing values correspond to those expected if the skip rule is correctly observed. The computationally time consuming multi-sampling approach was not used in the current research, and in some respects appears redundant given that it is already known that the data is not MCAR. Future revisions to the analysis reported here may consider the merits of this option.

On the basis of the foregoing discussion the EM algorithm was used to obtain maximum likelihood estimates of the model parameters and their standard errors for all WASC-Y items included in the analysis. The Yuan-Bentler scaled statistic was used as this is preferred over the Satorra-Bentler scaled statistic when data are both incomplete and non-normal (Yuan & Bentler, 2000). Of import, is that the normality assumption should not be expected to hold for the suicide subscale. Under the MAR assumption, standard errors were computed using observed not expected information as Schafer and Graham (2002) and Enders (in press) recommend. Note however, that model modification statistics were generated using the Fisher information matrix, as results to the LM test using the observed information matrix produced results that were too ambiguous to inform respecifications. When using the EM type of maximum likelihood the moment matrix was used as the data being modeled includes sample means and sample covariances (Byrne, 2006).

Convergent and Discriminant validity

Convergent validity requires that each pattern coefficient should have a non zero and statistically significant loading on only its target factor and that it should not have any statistically significant loadings on non-target factors. Discriminant validity can be assessed by constraining the correlation between two constructs to 1.00 and then using the chi-square difference test to compare the constrained and unconstrained models (Anderson & Gerbing, 1988). "A significant lower χ^2 value for the model in which the trait correlations are not constrained to unity would indicate that the traits are not perfectly correlated and that discriminant validity is achieved" (Bagozzi & Phillips, 1982, p. 476). The utility of the magnitude of this difference is less clear however, and would depend on the purpose for which the instrument is used. Note, each pair of correlations was tested independently, that is, not simultaneously with other constructs, and that an adjustment was made to the alpha level given that multiple χ^2 difference tests were made, alpha was set at [1 – (1 – alpha) x number of tests made] (Anderson & Gerbing, 1988).

Reliability and proportion of variance

Composite reliability for congeneric measures (Raykov, 1997), hereafter denoted using ϱ was used to estimate the internal consistency of each subscale for two reasons. First, ϱ does not consider correlated error terms as true variance and therefore this source of variance is excluded from the ϱ coefficient, and second, unlike Cronbach alpha, does not assume that the items are tau-equivalent. Unlike ϱ, in those instances where items are not tau-equivalent, Cronbach's alpha is a lower bound estimate. According to Nunnaly (1978) an internal consistency estimate larger than .70 for items of a single congeneric set is the minimum acceptable in ordinary research contexts. The use of ϱ with factors that have two or three indicators that are neither tau equivalent nor parallel has no utility as these factors are under identified when considered in isolation, that is, they have 0 or -1 degree of freedom. Therefore, in addition to Raykov's estimate ϱ, construct reliability will also be used where factors have two or three indicators. Construct reliability is "…defined in the classic sense, as the proportion of true variance relative to total variance (true plus error variance)" (Ullman, 2000, p. 715). The proportion of variance in the items that is accounted for by each factor will also be calculated (see Ullman, 2000, p. 716).

Model modification

The Lagrange Multiplier (LM) test was used to identify sources of misspecification in the model. The LM test statistic provides information about the "…absolute amount of χ^2 change that would result if parameters that were formerly fixed were free in a specified model" (Hoyle, 1995, p. 8-9). Typically parameters that are set to zero may have a large LM χ^2 statistic, suggesting that the restriction of the parameter may not be realistic in the population. Freeing the parameter may reduce the LM χ^2 statistic; however such decisions *must* be substantiated by sound theoretical rationale to avoid modelling sample specific variance. Type I errors can be reduced by again cross validating the respecified model on a second independent sample (MacCalum, 1993) although this is not possible in the current instance due to sample size restrictions. Misspecifications may relate to items that may better reflect different factors from the one specified, or they may correspond to correlated errors of measurement.

Correlated errors were not modelled in the current analysis. The emphasis rather was to highlight potential sources of misspecification with a view to inform future revision of the WASC-Y scales. Having said this, it behoves any good analyst to investigate if modelling any error covariance influences model parameters (Bagozzi, 1983) or if in constraining correlated errors to zero then biases error variances estimates (Alwin & Jackson, 1980).

Results

An assessment of the WASC-Y's construct validity proceeded under the assumptions of the common factor model (Lord & Novick, 1968). Four items within the WASC-Y were not consistent with the assumptions of this model and were excluded from the confirmatory factor analysis. These include items 15, 18, 21 and 22. These items are however, of demonstrated importance to any standard clinical assessment and should therefore not be ignored simply because they are not included in the analysis considered here. A consideration of their validity extends beyond the scope of this analysis. A confirmatory factor analysis of the covariance matrix using the original specifications outlined by Westerman (2002) tested the goodness-of-fit of the model to the sample data for the factors of depression, impulsivity, anxiety, suicide, and hopelessness. The fit of the model is reported in Table 8 below.

Table 1 CFA of the WASC-Y (n = 327)

Model	χ^2	Y-B$_{\chi}^2$	*df*	*Y-B* $_{\chi}^2$-*p* =	SRMR	*RMSEA	*RMSEA 90% CI	*CFI
Null model	3969.76	3169.65	465					
Model A	873.04	700.34	424	6.505E-16	0.049	0.045	.039,.050	0.922

Note: χ^2 = Chi Square; Y-B$_{\chi}^2$ = Yuan-Bentler scaled chi-square; *df* = degrees of freedom*; Y-B-p = probability for the* Y-B$_{\chi}^2$; SRMR = Standardised Root Mean Square Residual *RMSEA = Robust Root Mean Square Error of Approximation; *RMSEA 90% CI = 90%confidence interval for *RMSEA point estimate; CFI = Robust Comparative Fit Index

The SRMR and the *RMSEA were pleasing and the *RMSEA closeness of fit test was significant. The *CFI did not exceed Hu and Bentler's (1999) cutoff for a good fitting model, although a value of .922 is not unacceptable. Broadly these findings suggest that the measurement model can be considered to be an adequate fit to the sample data. The correlations between the WASC-Y factors are reported in Table 9. Factor correlations were all statistically significant at $p < .05$. The correlation between the two factors Depression and Anxiety was constrained to 1.00 and a chi-square difference test used to test the difference between the constrained and unconstrained models. The constrained model, Y-Bχ^2 (425, N = 327) = 739.971, p = 2.194E-19, and the difference between this model and the unconstrained model was statistically significant, $\Delta\chi^2 = 38.28$, $\Delta df = 1$, p = 6.129E-10. While these two factors show evidence of discriminant validity, the correlations between these factors was very high (.877, p <.05) - the factors are converging. Although not reported here, further chi-square difference tests between the remaining factor correlations demonstrated discriminate validity, although again, some correlations, for example, between Depression and Hopelessness, and Anxiety and Impulsivity appeared to suggest that these factors were converging. Future construct validation work of the WASC-Y may wish to consider the influence of higher order factors.

Table 2 WASC-Y Factor Correlations (n = 327)

	Item	1	2	3	4	5
1	Depression	-	.050	.040	.044	.027
2	Suicide	.713	-	.053	.052	.058
3	Hopelessness	.842	.769	-	.057	.055
4	Impulsivity	.805	.632	.737	-	.041
5	Anxiety	.877	.631	.745	.800	-

Note: Factor correlations are reported below the diagonal while their associated robust standard errors are reported above the diagonal.

The standardised estimates are reported in Table 10 below. All parameters were statistically significant ($p < .05$). Referring to Table 10, the R^2 value represents the proportion of variance in each item that is accounted for by its factor. The total amount of variance explained in the items by their respective factors is also displayed together with construct reliability estimates and Raykov's ϱ. Raykov's ϱ was calculated on those cases for which no missing data was observed. Readers should avoid making comparisons between ϱ and construct reliability as sample sizes differed with each estimate used.

Items 4, 11, 14, 17, 21 and 36 were not strong performers and users of the WASC-Y may wish to consider these items independently from the scales they reflect. These items do however have substantial theoretical and clinical relevance within standard cultural and clinical assessments.

Inspection of the Lagrange Multiplier (LM) test results were used to examine if any estimated parameters were incorrectly specified. Of interest was whether items may be loading on non-target factors as this would force us to reconceptualise the meaning of the WASC-Y scales. Error covariances may represent qualities of the instrument and at other times they may reflect characteristics of the respondents (Aish & Jöreskog, 1990). A poorly designed instrument for example may exhibit error covariances because of overlapping item content or ordering effects (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). However, error covariances may also be influenced by a tendency to respond in a social desirable way or as Warr (1990) suggests, may reflect a response tendency to acquiesce to items on the sole basis of their negative hedonic tone. They may also represent a small factor that has not been modelled (Byrne, 2006). These sources of error represent a systematic source of influence and this influence is of greater concern than mere random error.

Table 3 Standardised parameter estimates

Item Abbreviated label	Depression	Suicide	Hopelessness	Impulsivity	Anxiety	Error	R^2
1 I feel sad	.689					.725	.475
2 I'd rather be alone	.514					.858	.264
3 I can be happy one minute	.695					.719	.484
4 I like to sleep a lot	.316					.949	.100
5 I find it hard to pay attention	.592					.806	.351
6 I feel tired	.582					.813	.338
7 I find it hard to sleep	.551					.835	.303
9 I'm not really interested	.468					.884	.219
10 Better off without me	.656					.755	.431
11 I pick fights	.444					.896	.197
12 Ending my life		.867				.498	.752
13 Way to end life		.792				.611	.627
14 Easy to follow through		.410				.912	.168
16 I would try to end my life		.623				.782	.389
17 Would never do it		.268				.963	.072
19 Life is getting worse			.795			.607	.632
20 Life will not get any better			.694			.720	.481
21 Future			.198			.980	.039
24 Stupid things				.544		.839	.296
25 I find it hard to sit still				.728		.686	.530
26 I don't listen to reason				.763		.647	.582
27 I worry about lots of things					.617	.787	.381
28 I find it hard to breathe					.744	.668	.554
29 I feel dizzy					.608	.794	.370
30 I start to shake					.592	.806	.351
31 I feel sick in the guts					.667	.745	.444
32 Face gets all red and hot					.606	.796	.367
33 I get all sweaty					.548	.836	.300
34 I worry for no real reason					.639	.769	.409
35 I feel on edge					.709	.705	.503
36 I have bad dreams					.409	.913	.167
Raykov's ϱ	.815	.766	-	-	.847		
Construct reliability	.823	.740	.651	.644	.850		
Proportion of variance	28.09%	36.91%	45.63%	35.13%	33.54%		

The results to the LM tests revealed that the largest contribution of model misspecification was due to two error covariances associated with items 32 (When I worry my face gets all red and hot) and 11 (I pick fights with people for no reason) and items 4 (I like to sleep a lot) and 6 (I feel tired and have no energy). The LM χ^2 univariate increments associated with the error covariance of items 32 and 11 = 23.535, with a standardised parameter change value of .281. The LM χ^2 univariate increment associated with the error covariance of items 4 and 6 = 23.173, with a standardised parameter change value of .277. A third error covariance between item 19 (I feel like my life is getting worse and worse) and item 11 was associated with a univariate increment of 18.01, with a standardised parameter change value of -.289. Together these three error covariances accounted for an approximate overall drop in χ^2 for the model of 64.76.

The error covariance between items 4 and 6 may suggest that these two items represent a small factor, perhaps corresponding to fatigue or lethargy. When this factor was modelled an improvement in fit was observed however the factor did not perform strongly and future modifications to the WASC-Y may consider finding items 4 and 6, two or three new friends who could happily share a lethargy factor.

One method that can be used to estimate the practical significance of modelling these error covariances is to correlate the factor pattern coefficients and factor covariances between the model as reported in Table 10, with those estimated in a model where the error covariances have been specified. A correlation coefficient less than .90 would suggest that the error covariances are affecting the model parameters (Byrne, Shavelson & Muthén, 1989; Ullman, 2001). The correlations between the baseline and respecified model for the factor pattern coefficients, the factor covariances and the error variances were .997, .991 and 1.00 respectively. These results show that the error covariances were not affecting the model parameters to an extent that would preclude their specification (Bagozzi, 1983; Fornell, 1983). However, modelling these error covariances did not improve the fit of the model nor would have constraining them to zero have biased the error variance estimates (see Alwin & Jackson, 1980). As the remaining error covariances were neither statistically substantial nor theoretically reasonable no respecifications were considered appropriate. The final WASC-Y model remains as that specified in Table 10.

Self-esteem and Cultural Resilience

The WASC-Y has incorporated a nine-item subscale, which purports to provide a measure of 'self-esteem' for Aboriginal youth. The rationale for the inclusion of this subscale is based solely on the desire expressed from within the Aboriginal community during the consultation phase of this project to investigate the relevance of self-esteem as a predictor of mental ill health amongst Aboriginal youth. The self-esteem subscale (Items 37-45) includes the following items:

37. I worry about doing well at school

38. I think I am okay looking

39. I have lots of friends

40. People like me

41. I think I am a good person

42. I have something I am pretty good at (e.g. sports, hobbies)

43. My parents care about me

44. I like being Aboriginal

In the original validation study of Westerman (2002) the self-esteem factor was analysed as a one factor model which was then expanded to include items 45 to 53 (refer to items below). Two models were tested in the current analysis. A two factor model with one factor representing self esteem (items 37 to 44) and one factor representing protective factors (items 45 to 53). Following from the initial analysis by Westerman (2002) a one factor model corresponding to the cultural resilience subscale (items 37 to 53) subsuming all items was also tested to investigate whether they were able to represent a single "Cultural Resilience" subscale.

45. I speak my Aboriginal language

46. I am friends with whitefellas

47. I know a lot about my Aboriginal culture

48. I like going to school

49. When I feel upset, I can talk to someone (e.g. my parents/friends)

50. There is someone I know who I look up to and admire

51. When I feel upset, I can usually do something to make myself feel better

52. When people say racist things to me, I get really upset

53. People reckon I am pretty good at sports

Table 4 CFA of the WASC-Y Cultural Resilience Subscale (n = 293)

Model	χ^2	S-Bχ^2	*df*	*S-B-p*χ^2 *-p* =	SRMR	*RMSEA	*RMSEA 90% CI	*CFI
Model A								
Null model	1277.19	1140.20	120					
Model B I-factor	450.78	408.59	104	1.077E-37	0.080	0.100	.090; .110	0.701
Model C 2-factor	435.939	394.723	103	9.757E-36	0.081	0.098	.088; .109	0.714

Note: χ^2 = Chi Square; S-Bχ^2 = Satorra-Bentler scaled chi-square; *df* = degrees of freedom*; S-B-p*χ^2 *-p = probability value for S-*Bχ^2; SRMR = Standardised Root Mean Square Residual; *RMSEA = Robust Root Mean Square Error of Approximation; *RMSEA 90% CI = 90% confidence interval for *RMSEA point estimate; CFI = Robust Comparative Fit Index.

The goodness-of-fit of model B and C to the sample data was unacceptable. The *RMSEA closeness of fit test was statistically nonsignificant, in spite of adequate power (power = 100%), and the *CFI for the one and two factor models was clearly unacceptable.

Items 37 to 53 were subject to a maximum likelihood exploratory factor analysis using promax rotation. Multiple decision criteria were used to decide upon the most appropriate number of factors suitable for extraction. The decision criteria included the RMSEA closeness of fit test, examination of residuals, the ECVI and Horn's (1965) parallel analysis. For the purposes of running Horn's parallel analysis on the reduced correlation matrix, O'Connor's (2000) program was used to generate 5,000 normally distributed random data matrices. Following Glorfeld (1995) the eigenvalue corresponding to the 95th percentile was selected for each of the 16 latent roots. The values corresponding to each 95th percentile were then compared to the corresponding eigenvalues generated from the WASC-Y data. Results from this analysis clearly supported the extraction of eight factors. Two through eight factors were extracted, each time systematically varying the exponential power, *k*, from 2 to 6 to maximise simple structure. In each instance no clear simple structure emerged and the analysis was abandoned. In summary, the self-esteem and cultural resilience factors failed to demonstrate a minimally acceptable degree of construct validity. Based on the initial analysis by Westerman (2002) the cultural resilience factor emerged based on an unintended desire expressed within the community to learn more about what may cause youth who come from extraordinary risk, to not go on and develop mental health problems. It also emerged as a result of the self-esteem scale not performing well either clinically or statistically. Nonetheless, the items within the cultural resilience subscale are still vital to ongoing research in that it is essential to learn more about the concept of protective factors with Aboriginal youth. This study and Westerman (2002) represents the first time that protective factors (as a cultural resilience scale) has been explored statistically and empirically. It is vital that the elements which best predict or offer moderation of risk continue to be explored and determined. This is due to the number of Aboriginal youth at risk which are continuing to escalate, but also, importantly, the research which indicates that focuses on building up known protective factors offers a greater chance of reducing suicide than focusing on risk factors alone. Many of these factors have been incorporated within the cultural resilience subscale. Therefore, despite its poor performance statistically, there is a convincing clinical and theoretical argument of the relevance of this scale within a standard clinical assessment. If removed from the subscale score and analyzed independently from the cultural resilience factor that they have been specified to reflect their meaning should be based on their wording alone.

Qualifications

The reader is reminded that the factor analyses reported here is not strictly confirmatory. Cross-validation on a random and independent sample is necessary to avoid having generated a model that has capitalised on sample specific variance. Caution should be exercised when interpreting the results reported in the final estimation. It is unclear how robust these estimates may be when using EM with non-normal data under the MAR assumption. Further work is required before more definitive conclusions can be drawn although Enders (2001) does report limited evidence that standard errors are relatively accurate under the MAR assumption when using the Yuan-Bentler Scaled χ^2. Last, the high factor correlations may suggest that a second-order factor could be present. Future revisions to the analysis reported here may consider testing the fit of a second-oder measurement model.

Dependence of observations and multi-level models

An important caveat of the current analysis pertains to the assumption of independence and its possible violation. The WASC-Y data has been sampled from a variety of locations (both rural and urban) around Australia and this has included sampling children from seven Western Australian schools. Data of this type are hierarchically nested. Children are located within schools, and these schools are located within geographical regions. These regions may in turn be either urban or rural and may be located in numerous states around Australia. Of concern therefore, is that respondents may be more similar within schools or geographical regions than would be predicted on a pooled data basis. The assumption of independence requires that individuals within clusters will have no common characteristics or perceptions (Byrne, 2006) – with regard to the WASC-Y scores at least. Typically, one-way random effects intraclass correlations are used to assess the extent of the dependence of observations.

In hierarchically nested data sets the covariance and correlation between items is a function of the lowest level, for example, students within schools and also a function of the covariance among items between different clusters, for example, schools or geographical locations. The fit of a CFA to a single pooled covariance (as undertaken in the current analysis) that ignores clustering has two consequences. First, indices of fit, parameter estimates, standard errors and their associated tests of statistical significance may be biased (Muthén, 1997; Muthén & Satorra, 1995) and second, there is the possibility that the within and between level covariances may represent different measurement models. It may be of substantive interest to examine the fit of each model when these are broken down into within- and between level components. Unfortunately, considerable reworking of the WASC-Y data base is required before multi-level models can be applied. Moreover, multi-level models can be problematic to fit in practice and many hundreds of clusters may be required in order to ensure estimates have desirable asymptotic properties (Byrne, 2006). As more data is collected future validation work on the WASC-Y may consider the use of a multi-level confirmatory factor analysis to determine the degree of bias that may have resulted in analysing nested observations. Readers should also be mindful of this possibility when interpreting results to the numerous statistical significance testes reported in Chapter 9.

Psychometric equivalence across gender or Aboriginal cohorts

Of possible future interest is whether different ethnic groups (or indeed males and females) have a stable and consistent interpretation of the WASC-Y factor structure. Might children in more remote and rural locations, conceptualise depression and anxiety in different ways and if so would the items that reflect each latent construct carry the same meaning or interpretive weight? That is, do different groups score the WASC-Y with equivalence of form, item scaling and latent factor covariances (see Bollen, 1989). Non-equivalence of form would mean that the number of dimensions is not the same for all groups of respondents. For example, males may see depression as having two factors while females may see depression as having one factor. Non-equivalence of factor pattern loadings would suggest that the item content is perceived differently across groups, that is, there would be non-invariant item scaling. Said differently, the scale would be in a different metric thereby bringing into question score comparability across groups. This is not unlike trying to compare five inches with five centimetres. It is interesting to speculate if responses to the depression item 8 "Sometimes I feel so sad I cry", may carry different meaning for males and females. Crying may be associated with more severe depression among males than females (Schaeffer, 1988) and if so, this item may not have an equivalent factor pattern loading. Non-invariant factor covariances would suggest that the relations among the factor correlations are not the same across the two groups of respondents. If the WASC-Y is found to be differentially valid across genders or ethnic groups of children this would not necessarily mean that the instrument is invalid but it would bring into question the interchangeability (comparability) of scores for these populations. This possibility should be borne in mind when interpreting the results reported in Chapter 9. In summary, future consideration may be given to the interesting possibility that different groups of Aboriginals may conceptualise aspects of their psychological distress in very different ways.