International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 17(4): 232–240 (2008)
Published online 12 September 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.260

Extracting more information from behaviour checklists by using components of mean based scores

JOHN R. TAFFE,[1] BRUCE J. TONGE,[1] KYLIE M. GRAY,[1] STEWART L. EINFELD[2]

1 Centre for Developmental Psychiatry and Psychology, School of Psychology, Psychiatry and Psychological Medicine, Monash University, Melbourne, Australia
2 Brain and Mind Institute, University of Sydney, Sydney, Australia

Abstract
Sums of responses to behaviour checklist items are commonly used as outcome measures. We argue for the use of mean scores. For sets of responses registering absence and presence at different levels of intensity of behaviours we also show that mean scores may usefully be 'decomposed' into separate measures of the range and the intensity of problematic behaviours. These separate measures are the proportion of items positively endorsed and the 'intensity index' – the proportion of positive scores that are above one. We illustrate their use with primary outcome scores from the Developmental Behaviour Checklist (DBC) in the Australian Child to Adult Development Study. The low mean scores of young people with profound intellectual disability are shown to be a function of the narrow range of behaviours they display rather than of the level of intensity of these behaviours, which is relatively high. Change over time in mean scores is shown to be attributable to change in both the range and the intensity of behaviours as young people age in the study. We show how the technique of measuring these two separate strands contributing to mean scores may be applied to checklists with sets of responses longer than the zero, one, two of the DBC. Copyright © 2008 John Wiley & Sons, Ltd.

Key words: behaviour, checklist, mean, proportion, intensity

Introduction

Authors of behaviour checklists such as the Child Behaviour Checklist (CBCL) (Achenbach, 1991), the Developmental Behaviour Checklist (DBC) (Einfeld and Tonge, 1995) and the Aberrant Behaviour Checklist (ABC) (Aman et al., 1985) have generally recommended the use of total scores to measure constructs of interest, which may be based on the whole collection of items in the checklist, or on subsets of the items. The mathematical simplicity of total scores, or sums of item scores (SIS), is appealing. Adding item responses together is a natural way of creating a 'super-score' which measures a psychological construct of which individual checklist items measure aspects. Typically respondents are instructed to use positive whole numbers (1, 2, 3, . . .) to register that the behaviour in question is part of the repertoire of the subject, and 0 (zero) if this is not the case. This coding pattern ensures that in a sample of completed checklists individual item scores will be positively correlated with the SIS that measures the overall construct, which is a desirable property in a 'super-score'. Typically also the positive responses greater than one (2, 3, 4, . . .) register progressively greater than minimal degrees of intensity with which the respondent perceives that the behaviour is present in the subject, and consequently items receiving relatively many higher-intensity responses in a sample of completed checklists will contribute relatively more to the SIS than other items, and be more highly correlated with the SIS than other items. This also is a desirable property. Possibly because of these desirable properties, the case for using a SIS with a checklist to measure a psychological construct is usually taken for granted.

But the idea of using a SIS as a measure is not completely without drawbacks. One of these is that the measurement scale of a SIS is a function of the number of item responses that contribute to it. A SIS of 10 on a scale consisting of a dozen items may be relatively high, whereas on a scale consisting of 50 or 60 items it may be quite low. A SIS has no 'inbuilt' standard of comparison. Checklist users may handle this by referring to published norm-based percentile equivalents of each level of each SIS scale associated with their checklists, or by applying standardizing transforms to the SIS.

A practical drawback of SIS measures relates to their calculation in the presence of missing responses. An approach taken by some researchers is to record a missing value for the SIS if some threshold number of responses to its contributing items is not exceeded. The rationale for this procedure is related to the fact of the SIS scale being dependent on the number of items it comprises. The threshold is often set at some high percentage of the number of items on the scale.

An alternative 'super-score' is the mean item score (MIS), which is just the SIS divided by the number of items contributing to it, which for each subject is the number items in the scale for which responses have been provided. The MIS has a value which is within the range of scores defined by the coded responses of each individual item. So, for example, if the items are scored 0, 1 or 2 the MIS is a number on the (continuous) scale 0–2. The range of the response codes defines a standard background scale with which any MIS is compared. A MIS of 0.3 means that the per-item average of the zeros, ones and twos registered for the subject on the scale in question is 0.3, whether the scale has a dozen or 50 or 60 items. The meaning of the MIS is independent of the number of items in the scale. Whether 0.3 is a relatively high or a relatively low score on the scale (with reference to the sample or to norms) is a separate question.

This lack of dependence of the scale on which a MIS is measured on the number of items in the checklist means that the threshold number of items required to register a valid measure of the scale score may be lowered, or removed altogether, depending on the distribution in the sample of completed checklists of the number of missing items on the scale. A recent direct use of the lack of dependence of the scale of MIS on the number of items used in its calculation was made in the development of a short form of the DBC (Taffe et al., 2007).

Another advantage of using mean based rather than sum based scores arises when, as is usually the case, there are a number of subscores based on subsets of the checklist items. It is clear from the mean based scores which subscores are relatively high and low, compared to other scores for the same subject. Though this is less information than that provided by norm based 'subscore profiles', in that it is within-subject information, it is immediately available.

The use of mean based scores is to be recommended on the basis of these advantages alone. The purpose of this article, however, is to show how, for checklists with the coding structure described earlier, analyses based on MIS may be enriched by considering, in parallel with MIS, two components of the information it summarizes.

MIS, being a linear transform (in fact just a multiple) of SIS, contains exactly the same statistical information as SIS. This information comprises the number of checklist items checked positively about the subject and the levels of intensity at which they have been checked. The information in the checklist responses not captured by SIS or MIS consists of which items are checked, with which particular levels of intensity. It is clear that the same value of SIS or MIS may be arrived at in many different ways (except in the cases of the mathematical extremes of zero and the maximum possible MIS). A simple example is a SIS of 20, say, from a checklist with a set of response alternatives coded 0, 1, 2. It is easy to see that, for example, any 20 items each checked with the value 1, or any 10 items checked with 2, or any 18 items checked with 1 combined with any other single item checked with 2 (and so on) all give rise to this SIS, if all other items but these are checked with 0.

Some of these different ways in which the same SIS may arise involve different numbers of items checked positively, indicating different breadths of the behavioural basis on which the SIS or MIS depends. This aspect of the underpinning of the MIS may be separately summed up by the proportion of items (positively) checked (PIC). In the example earlier, if we suppose the whole checklist or scale has 80 items, a SIS of 20 could be accompanied by a PIC as high as 0.25

DOI: 10.1002/mpr

(20 items out of 80), indicating a broad behavioural base, or as low as 0.125 (10 items out of 80) indicating a narrow base.

PIC itself may be thought of as a mean based score (it is the mean that would result if all positive responses were coded as 1) measuring for a subject the breadth of the basis of the construct of which SIS and MIS are measures. Unlike these latter two, PIC does come with a natural standard of comparison. Being a proportion, it lies between 0 and 1, and 1 is the standard of comparison (as 100 is for percentages), even though in practice, if it relates to the whole checklist, a PIC of 1 will usually not be attained by any subject in a sample.

The information about the levels of intensity of positively checked items captured by SIS and MIS is missing from PIC. For checklists with responses coded 0, 1, 2 (such as the CBCL and the DBC) we may calculate the 'intensity index' (II) to capture this missed information. The intensity index is the proportion of the positively checked items which are scored 2. It can be shown that MIS, PIC and II are connected by the relationship MIS = PIC(1 + II), which confirms that PIC and II together contain all of the statistical information in MIS (which is the same as that in SIS). The relationship may be read 'the mean item score is the proportion of items checked positively increased by the proportion of twos among the positively checked items'. If every item of the checklist for a subject is scored either as 0 or 1 then II = 0 and MIS = PIC. If for another subject every item is scored either as 0 or 2, then II = 1 and MIS = 2PIC (the MIS is twice the PIC). These are the lower and upper bounds of MIS. Specifically, PIC ≤ MIS ≤ 2PIC, a fact that ensures that MIS and PIC are highly correlated (see Figure 1).

We illustrate some of the ways in which MIS and its components PIC and II can be useful with some analyses based on data in the Australian Child to Adult Development (ACAD) study (Einfeld and Tonge, 1996a, 1996b; Tonge and Einfeld, 2003; Einfeld et al., 2006), and discuss generalizations to checklists having longer sets of response alternatives.

Method and sample

The ACAD study has been well described elsewhere (Einfeld and Tonge, 1996a, 1996b; Tonge and Einfeld, 2003; Einfeld et al., 2006). Within the ACAD study the epidemiological subset is based on a near-complete identification of children and adolescents aged 4–18 years with intellectual disability in the moderate, severe or profound range and those with mild intellectual disability who used any health, education or welfare service in representative regions in New South Wales and Victoria.

The present analyses are based on all those in the epidemiological sample for whom a DBC was completed for at least the first of the four data waves and who were under the age of 19 years at the time of the first wave of data collection. Analyses involving change over time are by random effects regression, which takes account of the longitudinal structure of the data. The time-dependent outcomes MIS, PIC and II are modelled as linear functions of the time-dependent variable age and the non-time-dependent variables gender and level of intellectual disability. A person's age at the time of any data collection wave may be expressed as the sum of his/her average age over the study time points at which he/she contributed data and his/her deviations from this average at particular time points. This allows the unequal age differences between data collection points for different people to be taken into account. It also allows separate estimation of the effects of ageing in the study and differences between people in average age (a reflection of differing ages at entry to the study).

Results

The mean age of those in the ACAD epidemiological subset who were under 19 at wave one and for whom DBC checklists were completed was 11.7, 16.1, 19.1 and 23.0 at waves one, two, three and four, respectively. Standard deviations of age remained fairly constant (varying from 4.1 to 4.2).

Figure 1 shows the relationship between the MIS and the PIC for those in the epidemiological subset of the ACAD study at the time of the first data collection (wave one). The lower bound represented by the line MIS = PIC is attained by many subjects. These are the subjects for whom the respondents used only the responses coded 0 and 1 – never the 'intense' response coded 2. MIS attains its upper bound, twice PIC, for a small number of subjects. For these subjects the respondents have used only the codes 0 and 2 – never 1. For them any behaviour checked is at the intense level. The dashed line (MIS = 1.4PIC) shows the average amount (40%) by which MIS exceeds PIC for these subjects at this time. The graph confirms the expectation that MIS and PIC are highly correlated ($r = 0.96$, $p < 0.001$).

DOI: 10.1002/mpr

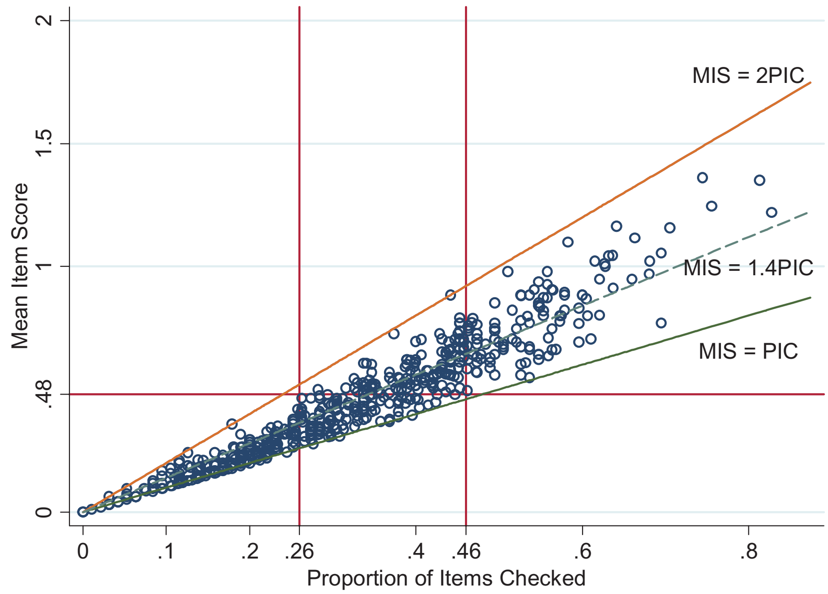

Figure 1. Relationship between MIS and PIC ACAD study, wave one, epidemiological subset.

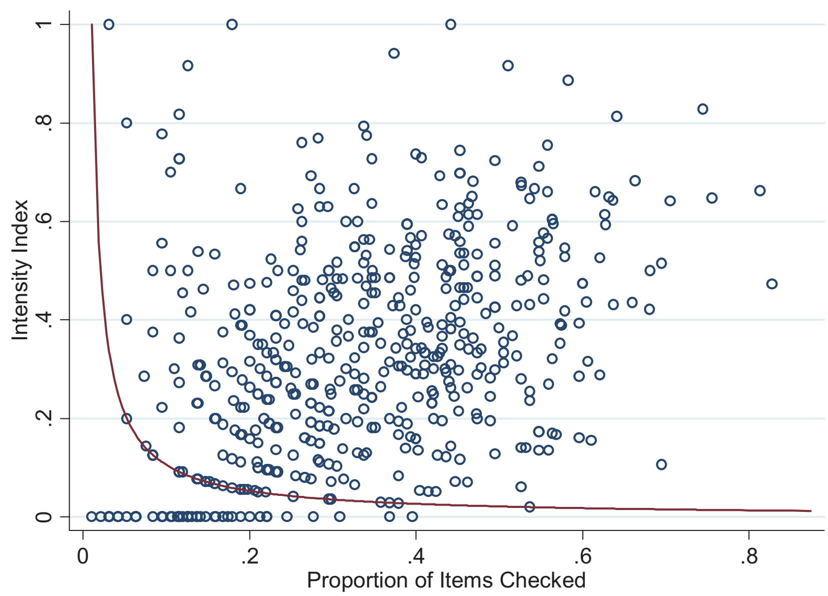

Figure 2. Relationship between II and PIC for the DBC in the ACAD study, wave one, epidemiological subset. The hyperbola shown joins points representing subjects for whom there is a single response at the intense level 2, the rest being ones or zeros.

The horizontal line at MIS = 0.48 corresponds to a SIS of 46, the cutoff for determining DBC-based psychological caseness (Einfeld and Tonge, 1995). The graph shows that subjects whose MIS is at about this level had anywhere between 26% and 46% of items checked positively about them. Those near the low end of this range had items checked at relatively high average intensity.

Figure 2 shows the relationship between the II and the PIC. There is evidence of some tendency for the

DOI: 10.1002/mpr

Table 1. Number of completed DBC checklists in the epidemiological subset[1] of the ACAD study by data wave, intellectual disability level and gender

Data wave	Intellectual disability level and gender							
	Mild		Moderate		Severe		Profound	
	Male	Female	Male	Female	Male	Female	Male	Female
1	109	86	127	95	66	44	12	11
2	85	62	109	78	58	40	8	10
3	83	64	98	76	51	39	7	7
4	82	65	96	78	54	34	6	6

[1] Restricted to those aged under 19 years at wave one.

plotted points to be higher at the right than at the left (that is, for those with higher PIC to have higher levels of the II), and thus of some positive association. In fact for this relationship $r = 0.40$ ($p < 0.001$). An interesting pattern is discernible at the lower left side of the plot. There are several collections of points lying along rectangular hyperbolas (the graph of one of which appears in Figure 2) above a horizontal row at level zero of the II. The horizontal row represents those subjects about whom respondents used no twos, only zeros and ones – the same subjects whose points lie along the line MIS = PIC in Figure 1. The points lying along the hyperbola whose graph is drawn represent subjects for whom the respondents used a single 2, with the rest of the responses zeros or ones. The highest of the plotted points lying on this curve represents a subject whose SIS of six results from four items scored 1, one item scored 2 and the rest scored 0. For this subject, MIS = SIS/95 = 6/95 = 0.06, PIC = number of items checked/95 = 5/95 = 0.05 and II = number of items checked with 2/number of items checked = 1/5 = 0.2, the height of the plotted point. [The right-most plotted point lying on this graph represents a subject who has 51 of the 95 items checked positively (SIS = 52, PIC = 0.54 and II = 0.02).] About 12 such hyperbolas, representing different small numbers of 'intense level' responses, can be detected with the naked eye. They are not of theoretical importance in the present discussion, but we thought we should explain the reason for their existence.

A longitudinal analysis

We now present a longitudinal analysis based on 1846 completed DBC checklists. Table 1 gives a breakdown of the numbers of checklists by data wave, level of intellectual disability and gender. The overall gender ratio of study subjects is 57 : 43 male to female, and the intellectual disability levels are in the ratios 34 : 41 : 21 : 4 for mild, moderate, severe and profound, respectively. The ratios of numbers of completed checklists in successive waves are 100 : 82 : 77 : 77 for waves one, two, three and four, respectively.

A consistent pattern for the three measures MIS, PIC and II emerges from Table 2. All have similar levels among those with mild and moderate intellectual disability and slightly higher levels among those with severe intellectual disability. MIS and PIC have lower levels among those with profound intellectual disability. There is no evidence of any consistent gender based difference. Each measure shows a reasonably consistent pattern of slight decrease in mean over time, from wave one to wave four.

For PIC we see that except for those with profound intellectual disability, about a third, on average, of the DBC items are given positive responses, regardless of gender and intellectual disability level. From the information on II we see that, on average, 30% or fewer of the positive responses are at the 'intense' level (2) for those with mild or moderate intellectual disability, but the percentage is generally higher than 30% on average for those with severe or profound intellectual disability.

[The mean of MIS can be approximately calculated as mean PIC(1 + mean II) + 0.015 (for example 0.48 = approx 0.36(1.29) + 0.015, using figures in the top left cells in Table 2 for the three measures). The 0.015 is the overall estimate from the DBC responses used in this study of the covariance of PIC and II.]

Table 2. Means of three DBC summary measures in the epidemiological subset[1] of the ACAD study by data wave, intellectual disability level and gender

		Intellectual disability level and gender							
Measure	Data wave	Mild		Moderate		Severe		Profound	
		Male	Female	Male	Female	Male	Female	Male	Female
MIS	1	0.48	0.45	0.44	0.45	0.50	0.48	0.30	0.24
	2	0.45	0.40	0.44	0.43	0.50	0.43	0.29	0.34
	3	0.38	0.44	0.41	0.39	0.45	0.48	0.25	0.35
	4	0.38	0.36	0.36	0.38	0.43	0.47	0.19	0.33
PIC	1	0.36	0.33	0.33	0.33	0.35	0.33	0.21	0.16
	2	0.34	0.31	0.33	0.33	0.36	0.31	0.22	0.24
	3	0.30	0.34	0.31	0.31	0.33	0.34	0.19	0.27
	4	0.29	0.29	0.28	0.30	0.32	0.33	0.14	0.25
II	1	0.29	0.30	0.30	0.31	0.41	0.42	0.45	0.48
	2	0.27	0.24	0.27	0.27	0.36	0.33	0.32	0.37
	3	0.26	0.24	0.25	0.22	0.33	0.36	0.22	0.31
	4	0.26	0.19	0.22	0.21	0.33	0.40	0.44	0.28

[1] Restricted to those aged under 19 years at wave one.
Note: MIS, mean item score; PIC, proportion of items checked; II, intensity index.

Table 3. Longitudinal regressions of MIS, PIC and II on age gender and intellectual disability level in the epidemiological subset[1] of the ACAD study

		MIS	PIC	II
Average age		−0.005*	−0.003*	−0.002
Deviation from average age		−0.007***	−0.004***	−0.007***
Female		0.000	0.001	−0.009
Intellectual disability level				
(Reference group: mild)	Moderate	−0.003	−0.004	0.002
	Severe	0.045	0.009	0.110***
	Profound	−0.144***	−0.123***	0.141***
(constant)		0.499***	0.376***	0.288***

[1] Restricted to those aged under 19 years at wave one.
Note: MIS, mean item score; PIC, proportion of items checked; II, intensity index.
$*p < 0.05$; $***p < 0.001$.

The regression analyses for MIS and for PIC (Table 3) tell the same story: on average the level of problem behaviour as measured by either of these scores declines significantly by a very small amount for each year of ageing during the study, is negatively associated with age at entry to the study, is not related to gender but is considerably lower for those with profound intellectual disability. For PIC, which is a proportion, we can refer to the 0–1 scale against which proportions are gauged and say that its average rate of decrease is 0.4 percentage points (of this scale) per year.

Int. J. Methods Psychiatr. Res. 17(4): 232–240 (2008)

DOI: 10.1002/mpr

From Table 3 it can also be seen that on average II decreases significantly but slowly with ageing in the study, is not associated with age at entry to the study nor with gender but is higher by a margin of 11% among those with severe intellectual disability and by a margin of 14% among those with profound intellectual disability, compared to the reference group with mild intellectual disability, after age and gender are controlled.

Both the extent of problematic behaviours, measured by PIC, and their level of intensity, measured by II, have been shown to decline somewhat with ageing in the ACAD study. MIS, a directly dependent combination of these two measures, is constrained to do the same. That is, the mean (and the sum) of item scores declines because on average fewer troublesome behaviours are reported, and the intensity at which they are reported declines as well, at a similar rate. Further, those with severe and profound levels of intellectual disability have troublesome behaviours reported at higher intensity than those with mild or moderate levels of intellectual disability, when age and gender are controlled. While those with severe intellectual disability do not differ on MIS or PIC from those with mild intellectual disability, they do have behaviours reported at higher intensity.

Discussion

Sums of item scores are a conceptually simple way of summarizing some of the information reported by survey respondents about which behaviours are displayed at what levels of intensity by subjects they are reporting on. Per item means are statistically equivalent to sums, in that they contain the same information, but have the advantages that they are reported on the scale defined by the set of responses, regardless of how many items they contain, and that in general, restrictions on the minimum number of items responded to need not be applied. But both sums and means of item scores contain, and hide, two kinds of information. These are the number of items endorsed, which measures the extent of the basis of the problematic behaviour characteristic of the subject, and the level of intensity, as perceived by the respondent, of these behaviours.

We have shown that it is easy to examine both these aspects of the information collected, by considering both the proportion of items positively checked, which measures the extent of the basis of the problematic behaviour, and the II, which measures the degree of intensity at which the behaviours are manifest.

The example we used was of a checklist (the DBC) which uses a three-point scale of responses coded 0, 1, 2. With longer sets of response alternatives, such as the four-point scale (0, 1, 2, 3) used by the ABC (Aman et al., 1985) and the Nisonger Child Behaviour Rating Form (Lecavalier et al., 2004) the II we have defined may be used if the fourth point, 3, is regarded as a 2 (that is, if the set of responses is collapsed to that of the DBC and the CBCL). This would give a 'first-order approximation' to the intensity level as collected. But it is also easy to extend the II definition to allow for two levels of intensity above the basic 1 which registers that the behaviour is present. We may define Intensity Step 1 as the proportion of twos and threes (considered as equivalent) among the positive responses (the ones, twos and threes). This is analogous to the II we have defined for checklists with responses coded 0, 1, 2, and is the proportion of 'more intense than just there' responses among those checked positively. Intensity Step 2 may be defined as the proportion of threes among the twos and threes, or the proportion of 'very intense' among the 'intense' responses. With these definitions it can be shown that MIS = PIC(1 + IS1(1 + IS2)), showing that all of the information contained in MIS (and SIS) is captured by the three measures PIC, IS1 and IS2. And of course, if longer sets of responses with the same construction are used, extra intensity measures may be similarly defined, and the identity connecting MIS with PIC and the intensity measures extended.

The upper and lower bounds on the area in which plotted points may lie, depicted in Figure 1, show that for scales with responses coded 0, 1, 2, MIS and PIC can always be expected to show very strong linear dependence. In the case of the DBC responses this dependence is so strong that MIS and PIC may be thought of as interchangeable measures of the extent of problematic behaviour, a fact borne out by the longitudinal analyses. But even when this is the situation, the same analyses show that considering the separate information summarized in the II can deliver more than any analysis based only on MIS or SIS.

For the DBC we have found that the known gradual decline in the total behaviour score with ageing in the study (Einfeld et al., 2006) cannot be attributed just to a gradual lowering of the intensity with which parents and carers respond to problematic behaviours, nor just

DOI: 10.1002/mpr

to a reduction in the range of problematic behaviours the young people produce, but is due to a combination of these two sources. Why this should be is still a matter for conjecture, but we do know that both these change phenomena are occurring. We could not have discovered this from analysing sums or means of item responses alone. We have found also that low total and mean behaviour scores for those with profound intellectual disability are due to the restricted range of behaviours such people produce, and that the narrowness of this range of behaviours outweighs (in mean and total scores) the markedly high levels of intensity at which respondents assess their problematic behaviours. We have found further that a relatively high level of intensity of response is characteristic also of those with severe intellectual disability (as well as those with profound intellectual disability). This information also is not available from a study of total or mean scores.

For checklists with a four-point response scale the searchlight-shaped area in which the points in the plot of MIS against PIC lie (shown for a three-point scale in Figure 1) widens (the linear bounds on MIS are PIC and 3PIC), so the correlation between MIS and PIC can be expected to be lower than it is for checklists with three point response scales.

The plot of II against PIC (Figure 2) and the low correlation between these measures show that the II is essentially different information from PIC, at least in the context of the DBC-based information in the ACAD study, which is something we also know from the new information yielded by our analysis of II.

In the fourth data wave of the ACAD study two slightly different forms of the DBC were used. Three quarters of the subjects, who had by then passed their 19th birthdays, were reported on via the DBC-A (Mohr et al., 2005), a variant for adults of the original DBC-P, which was designed for 4–18 year olds. The DBC-P, of 95 substantial items, and the DBC-A, of 106, share 94 items (though nine of these are worded slightly differently in the two forms). The deletion concerns mixing with children of a different age group (not appropriate in an adult context) and the additions tap adult-specific depressive and psychotic symptoms, and problems with legal and illegal drugs. We have taken advantage of the fact that scales of measurement of MIS, PIC and II are independent of the number of items in the checklist (which is not the case with SIS), and calculated these three measures from whichever of the two forms was used for the subject at the fourth data wave. In doing this we do not claim that the DBC-P and the DBC-A are identical instruments. We claim only that the SIS and MIS calculated from them are essentially the same measure of the same psychological construct, psychopathology, as evidenced by the level of problematic behaviour. In support of this, an earlier SIS-based analysis of DBC responses in the ACAD study (Einfeld et al., 2006), in which the DBC-A responses only of the items common to the two forms were used, found an average decline with ageing of about 1 SIS point per year, equivalent to the approximate 0.7 point SIS per year of the present results. This is good consistency in studies with different analysis approaches and a slight difference in the way levels of intellectual disability are accounted for (but both taking initial age, ageing in the study, gender and level of intellectual disability into account).

We have shown a simple way to separate into two strands the information from behaviour checklists captured by sums of scores of items covering the whole checklists or subsets of them. These two strands are the extent of the behavioural basis of the trait being measured and the intensity at which the trait is perceived by the respondent. The SIS contains both of these kinds of information but reveals neither. The two kinds of information may often be of interest in their own right, but it will be most useful in general to consider them in parallel with the MIS. The MIS itself is a simple transformation of the usually reported SIS, but has the advantage over the sum that, whether it is the overall or a subscale measure, it is reported on the scale defined by the codes allocated to the set of response alternatives of the checklist, which enables immediate within-person score comparisons.

Using these mean based scores is not incompatible with using standardized and norm referenced scores based on sums (totals) of items. These score transforms can themselves easily be expressed in mean rather than sum terms. Clinicians and researchers alike will find it beneficial to consider MISs, PICs and IIs of the type we introduce here.

Acknowledgements

The authors acknowledge helpful comments received in correspondence with Scott Hofer, Andrea Piccinin, Daniel Bontempo, Andrew Mackinnon, John Carlin, Michael Aman and Thomas Achenbach. The authors acknowledge continuing funding for the Australian Child to Adult Development study from the Australian National Health and Medical

Research Council and support from the National Institute of Mental Health (R01MH61809).

Declaration of Interests

The authors have no competing interests.

References

Achenbach TM (1991). Manual for the Child Behavior Checklist/4–18 and 1991 Profile. Burlington, VT: University of Vermont, Department of Psychiatry.

Aman MG, Singh NN, Stewart AW, Field CJ (1985). The Aberrant Behavior Checklist: a behavior rating scale for the assessment of treatment effects. Am J Ment Defic 89(5): 485–491.

Einfeld SL, Tonge BJ (1995). The Developmental Behaviour Checklist: the development and validation of an instrument to assess behavioural and emotional disturbance in children and adolescents with mental retardation. J Autism Dev Disord 25(2): 81–104.

Einfeld SL, Tonge BJ (1996a). Population prevalence of psychopathology in children and adolescents with intellectual disability: I. Rationale and methods. J Intellect Disabil Res 40(2): 91–98.

Einfeld SL, Tonge BJ (1996b). Population prevalence of psychopathology in children and adolescents with intellectual disability: II. Epidemiological findings. J Intellect Disabil Res 40(2): 99–109.

Einfeld SL, Piccinin A, Mackinnon A et al. (2006). Psychopathology in young people with intellectual disability. JAMA 296(16): 1981–1989.

Lecavalier L, Aman MG, Hammer D et al. (2004). Factor analysis of the Nisonger Child Behavior Rating Form in children with autism spectrum disorders. J Autism Dev Disord 34(6): 709–721.

Mohr C, Tonge BJ, Einfeld SL (2005). The development of a new measure for the assessment of psychopathology in adults with intellectual disability. J Intellect Disabil Res 49(7): 469–480.

Taffe JR, Gray KM, Einfeld SL et al. (2007). A short form of the Developmental Behaviour Checklist. Am J Ment Retard 112(1): 31–39.

Tonge BJ, Einfeld SL (2003). Psychopathology and intellectual disability: the Australian Child to Adult longitudinal study. In Glidden LM (ed.) International Review of Research in Mental Retardation. San Diego, CA: Academic Press. pp 61–91.

Correspondence: Dr John R Taffe, Centre for Developmental Psychiatry and Psychology, Child and Adolescent Mental Health Service, Monash Medical Centre, 246 Clayton Road, Clayton VIC 3168, Australia.
Telephone: (+61) 3 9594 1381
Fax: (+61) 3 9594 6333
Email: john.taffe@med.monash.edu.au