Vulnerable Children and Youth Studies, 2007, 2, 2, 127-141

Reliability of Self-Report of Health in Juvenile Offenders

Dianna T Kenny[a] and Jennifer Grant[b]

Abstract

The aim of the present study was to investigate the accuracy of self-reports of juvenile offenders on physical factors (e.g., sleep difficulties, weight related behaviors and weight perceptions), health risk behaviors (e.g., alcohol use), trauma history (e.g., physical and sexual abuse) and psychological factors (e.g., anxiety, suicidal and self-harm behaviors). Self-reports obtained via a Health Questionnaire from 242 incarcerated juvenile offenders were compared with standardized measures (Body Mass Index, *Adolescent Psychopathology Scale* and *Child Trauma Questionnaire*) to investigate the reliability (via construct validity) and veracity of their self-report. Using kappa estimates and receiver operating characteristic curves, results generally showed high agreement across measures, suggesting that self-report questions from the health survey could all be used reliably. The degree of accuracy indicated that young offenders are as reliable as clinical and community samples of adolescents in their self-report. These findings have implications for routine assessments and practice evaluations that rely on self-report as the method of data collection and as the basis for clinical formulation and treatment planning.

Keywords: juvenile offenders, self-report, reliability, health

Correspondence concerning this article should be addressed to
Professor Dianna T. Kenny, Faculty of health Sciences C42, University of Sydney
PO Box 170, Lidcombe NSW 1825
Ph: +61-2-9351-9644; Fax: +61-2-9351-9540
Email: d.kenny@usyd.edu.au

[a]University of Sydney, Australia

[b]University of New South Wales, Australia

Introduction

Provided that it is reliable, self-report has many advantages; it is cost-effective, less time consuming, and non-invasive than other forms of data gathering in health settings (Waters et al. 2003). Health professionals must rely on self-reports of many health-related behaviors such as alcohol and substance use because objective indicators are often unable to provide information on specific aspects of user behavior such as history, frequency and severity over extended time periods (Carroll 1995). Consequently, self-report of health behaviors is a widely used method of data collection. However evidence regarding the accuracy of self-reports in juvenile forensic populations is limited, and of the available evidence, results are inconsistent. Hence the accuracy of self-report on a range of common health related behaviors and concerns - sleep difficulties; alcohol use; physical and sexual abuse; perceptions of weight and weight related behaviors; anxiety; and suicide and self harm - were assessed in a juvenile forensic population and compared with accuracy of self-report in community and clinical samples of adolescents, which has generally been reported to be acceptable to highly reliable.

For example, adolescents frequently report sleep difficulties (Ohida et al. 2004). When compared with objective measures such as diary reports and actigraphy (ie sleep activity-based monitoring) estimates, self-reports of sleep patterns of adolescent school students were generally accurate and reliable, with self-reported sleep estimates lying within five minutes of the results obtained from the more objective estimates (Wolfson et al. 2003). However, evidence concerning the accuracy of self-reported sleep behaviors from the juvenile offender population is lacking.

Reliability of self-report of alcohol use by adolescents from community mental health samples, psychiatric settings, and the general population is generally high when compared with objective measures such as urinalysis results, with relatively low rates of under-reporting of alcohol use recorded (Del Boca et al. 2003; Johnson et al. 2004; Lintonen et al. 2004; Shillington et al. 2000; Williams et al. 2005). Williams et al. (2005) study of the accuracy of adolescents self-report from an outpatient sample indicated that 93% of those who reported no alcohol use tested negative for ethanol by urinalysis testing. However, results are mixed for young offenders. Some researchers have reported poor agreement between self-report of drug use, with fewer than 50% of the sample of 3,086 juvenile arrestees reports of recent substance use in agreement with urinalysis results (Fendrich et al. 1994). Although, young offenders appear to self-report substance use more accurately than adult offenders, shown by 65% agreement between self-report and urinalysis results compared to 55% for an adult offender sample (Yacoubian et al. 2003).

Victim self-reports is an important source of information regarding most forms of maltreatment (Lipschitz et al. 1999). Self-reports of both physical and sexual abuse in the broad adolescent population, high-risk samples (Lipschitz et al. 1999; Paivio 2001) and adolescent psychiatric in-patients are generally reliable and valid when compared with hospital records and clinician interviews (Winegar et al. 1999; Wekerle et al. 2001). Lipschitz et al (1999) found an adolescent inpatient sample to be generally consistent and accurate reporters of physical and sexual abuse, represented by 86% and 71% respective agreement across the measures used. Again, little is known about the accuracy of these reports from a juvenile offender population, a high-risk population for experiencing maltreatment.

Perceptions of weight and body shape are common concerns amongst adolescents (Field et al. 2004). However, there is limited evidence for the validity of self-reports of eating disordered behaviors in adolescent populations, most of which focus on females. To date, little evidence pertaining to the accuracy of self-reported perceptions of weight and associated behaviors by a juvenile offender sample has been established. Studies from the general adolescent population and inpatient samples show consistent support for the accuracy of adolescent self-report of eating disordered behaviors when compared with investigator-based interview, particularly for binge eating (Field et al. 2004), purging habits (Fairburn et al. 1994), weight concerns, compensatory behaviors and eating restraint (Decaluwe et al. 2004; Passi et al. 2003).

Adolescent inpatients' self-reports of anxiety-related behaviors have been validated against

structured interviews and observer ratings with results indicating agreement of greater than 70% between the self-reports and structured psychiatric interviews (Gadow et al. 2002). Weaker designs have assessed agreement using parental or proxy reports as the alternative measure (Ferdinand et al. 2004) but it is difficult to draw conclusions from such designs as parental reports have not been established as objective and reliable. In addition, adolescent health studies may be limited in their generalizability to other adolescent populations as recruiting solely through health services is unlikely to provide accurate information about "at risk" adolescents who do not come into contact with health agencies (Waters et al. 2001). Currently, there are no studies of the accuracy of self-reported psychopathology by juvenile offenders.

Little is known about the accuracy of adolescents' self-reports of suicide and self-harm ideation and behaviors. Some studies on clinical and outpatient samples show that adolescents provide valid self-reports of this behavior when compared with structured interviews and clinical records (Safer 1997) although others have reported a tendency for adolescents to minimize or deny such behaviors. In one study, 50% of the outpatient sample provided inaccurate accounts of suicidal behavior when compared with clinical measures (Velting et al. 1998). However, anonymity and assurances of confidentiality in research on adolescent suicide and self-harm improve accuracy of reporting (Safer 1997; Velting et al 1998). There are no studies examining the reliability of juvenile offender self-reports of suicidal ideation or behavior.

Recent research has indicated that juvenile offenders may be more likely than other adolescent samples to reveal alcohol use and childhood abuse as this population may not perceive as high a social cost to reporting these behaviors (Johnson et al. 2000; Percy et al. 2005). Demographic variables (e.g., age, gender and education), interview factors (e.g., setting, and question content and form) and interviewee characteristics (e.g., cognitive abilities, and social desirability or response biases) also have the potential to influence the accuracy of self-report, leading to under- or overestimations of behaviors (Del Boca et al. 2003; Johnson et al. 2004; Passi et al. 2003; Santelli et al. 2002; Tanofsky-Kraff , Morgan et al. 2003).

Further, there is no "gold standard" for determining the accuracy of adolescents' self-reports on a variety of health related behaviors due to the potential fallibility of all available measures (Tanofsky-Kraff, Yanovski et al. 2005). Nonetheless, young offenders typically present with numerous health concerns and high risk behaviors (Langhinrichsen-Rohling et al. 2004). Thus, evidence is needed regarding the accuracy of their self-report of health to support the reliance on this source of information in their health care.

The Present Study

Self-reports of juvenile offenders were obtained via a Health Questionnaire (HQ) containing questions worded in a manner consistent with those asked in a routine health consultation. To assess the construct validity of specific sections of the questionnaire, responses from the HQ were compared with reliable information collected using standardised tools [Body Mass Index (BMI), *Adolescent Psychopathology Scale* (APS), and *Childhood Trauma Questionnaire* (CTQ)]. This study investigated the extent to which HQ responses predicted the BMI, APS, and CTQ standardised outcomes.

Method

Ethical approval for the study was obtained from the ethics committees of the NSW Department of Juvenile Justice, JusticeHealth and the Aboriginal Medical Health and Research Council.

Participants

Participants were young offenders serving custodial orders with the New South Wales Department of Juvenile Justice who volunteered to take part in *The Young People in Custody Health Survey* (YPiCHS) (NSW Department of Juvenile Justice (2003).

All consenting young people in custody in NSW who were serving remand or periods of control in the study period were eligible. Either the severity of the offence or the offenders' criminal history in terms of commission of past offences

led to incarceration in NSW juvenile justice detention centres.

At the time of the study there were nine such centres in NSW, five in urban regions and four in rural regions. There were 319 young people eligible to participate and 242 actual participants (223 males and 19 females) (75.86%). Eighteen percent (18.1%) participants were age 14-15 years, 59.1% were age 16-17 years and 21.8% were 18 years or older. Ethnic group membership was as follows: n=102 (42%) were from English speaking backgrounds; n=102 (42%) were from Aboriginal and Torres Strait Islander background; and (n=40; 16%) were from culturally and linguistically diverse backgrounds. Offences leading to incarceration were robbery (28%); break and enter (21%); assault (17%); car and other theft (10%); sexual assault (7%); aggravated assault (6%); homicide (5%); and other (eg traffic offences, stalking, kidnap) (6%). Thirty-five percent (35%) received custodial sentences of less than six months; 29% were incarcerated for between 6-12 months; 20% for 1-2 years; 15% for 2-5 years; and 1% for more than 5 years.

Measures

Health Questionnaire (HQ)

The HQ contains 367 questions divided into 32 sections, family history, parental characteristics, living arrangements, educational background, employment history, self-reported health status and health behaviors, including health education, physical activity, sun protection, nutrition, disability, recent symptoms, medication, injury, and health service utilization (including treatment for alcohol and substance abuse); and risk behaviors, including drug and alcohol use, sexual health, smoking, gambling, tattooing and body piercing. Index questions on the health issues selected for assessment were extracted from this questionnaire for comparison with results from the standardized measures described below.

Child Trauma Questionnaire (CTQ)

The CTQ (Bernstein et al. 1998) is a 28-item self-report inventory that provides brief, reliable, and valid screening for histories of emotional, physical and sexual abuse and emotional and physical neglect. Item scores are summed to produce the scale total score, (range 5 to 25); the higher the score, the greater the severity of maltreatment. There are four levels of maltreatment for each type of trauma: none (minimal); low (to moderate); moderate (to severe); and severe (to extreme). The higher the score is, the greater the severity of abuse for that scale. There are four categories of severity for each trauma type: None (minimal); Low (to Moderate); Moderate (to Severe); and Severe (to Extreme).

The cut-off scores for each scale are as follows:

Level of abuse	Emotional Abuse	Physical Abuse	Sexual Abuse	Emotional Neglect	Physical Neglect
No	8	7	5	9	7
Low	12	9	7	14	9
Medium	15	12	12	17	12
High	16+	13+	13+	18+	13+

There is also a minimization and denial scale scored either none (0) or possible (1 to 3). Internal consistency is in the satisfactory to excellent range (.66 to .92), with the total scale achieving a Cronbach's alpha of 0.95. Test-retest reliabilities were high (.79 to .86); and construct validity is generally robust, with psychiatrically referred groups reporting higher levels of abuse and neglect than non clinical samples (Berstein et al. 1998; Strand et al. 2005).

Adolescent Psychopathology Scale (APS)

The APS (Reynolds 1998) is a reliable and valid measure of psychological and psychiatric symptoms warranting possible referral or intervention. The 40 APS scales are based on DSM-IV criteria and are organized according to clinical disorders (20 scales), personality disorders (5 scales), psychosocial problems (11 scales) and response style indicators (4 scales). The APS has a mean T-score of 50 (SD = 10), and scores are categorized into five symptom classifications: no symptoms (below 50T), sub-clinical (50T–59T), mild (60T–69T), moderate (70T–79T) and severe (80T and above). Scales with T-scores in the 65T-69T range should be examined for psychopathology that is clinically significant, while scores in the severe range (80T) represent significant psychological problems.

Procedure

The study met design requirements to measure construct validity (Peat et al. 2001). The conditions under which the assessments were made were identical. Interviewers administering the HQ were blind to results from the standardised outcomes, and the assessments were undertaken at the same time to avoid time effects on the various measures. In addition, the HQ and the standardised outcomes were measured independently but in consistent circumstances. To assess the construct validity, responses from the HQ were compared with reliable self-report information collected using the standardized tools (BMI, CTQ and APS), accepted as reliable benchmarks from which to make these comparisons. The following methods were used for assessing agreement between HQ responses and the standardized outcomes, and for predicting standardized outcomes using the HQ.

Categorical HQ Responses Compared to Categorical Standardised Outcomes

The extent to which categorical responses agree was assessed by constructing contingency tables of HQ responses versus standardised outcome responses and then by measuring agreement between the two methods using the kappa statistic and the proportion of responses in agreement. For data with three or more possible responses and for ordered categorical data, weighted kappa was used. It is not possible to compare kappa values between questions. Kappa is an estimate of the proportion in agreement, in excess of the agreement that would occur by chance, between the two types of data collection methods. A value of one indicates perfect agreement and a value of zero indicates no agreement. In general, values less than 0.40 indicate poor agreement, values between 0.41 and 0.60 indicate moderate agreement, values between 0.61 and 0.80 indicate good agreement, and above 0.81 indicate very good agreement (Altman 1996). Kappa increases as the proportion of negative and positive responses become more equal for each response. Thus it was important to consider both kappa and the proportion in agreement in assessing which questions were most reliable, that is, having the highest construct validity. The strength of the association between categorical variables was measured using a continuity corrected chi-square test for 2x2 contingency tables and Pearson's chi square test for larger tables.

Categorical HQ Responses Compared to Continuous Standardised Outcomes

Where the standardised outcome is measured on a continuous scale, the ability of the standardised outcome to predict a categorical HQ response was assessed using receiver operating characteristic (ROC) curves. For each measurement, a cut-off value that had the greatest discriminatory value for delineating a positive from negative HQ response was calculated (Altman et al. 1994). Associations were also investigated using means plots with 95% confidence intervals and one-way analysis of variance at alpha level .05. All statistical analyses were conducted using the Statistical Package for Social Sciences, Version 12 (SPSS 2003).

Results

Tests of Assumptions for Analysis of Variance

Homogeneity of variance was satisfied for all analyses. Assumptions of normality were generally supported, except for a small number of minor violations of normality. Because tests of normality are conservative and there were no

outlying or extreme values that would have tended to inflate between-group differences, a parametric analysis of variance was used.

Sleeping difficulties

The sleep question "Do you have trouble sleeping?" was validated against the APS sleep disorder scale which was available both as a continuously distributed t-score and as a three-level categorical variable. Of the 227 responses to this question, 131(57.7%) indicated that they did not have trouble sleeping and 96 (42.3%) indicated that they did.

The agreement between the question "Do you have trouble sleeping?" and the APS t-score was examined using a ROC curve. The APS t-score was a significant predictor of the response to the question with a high area under the curve of 0.78 (95% CI 0.72, 0.84). The optimum cut-off t-score value for predicting trouble sleeping was 54.5 which gave a moderate true positive rate of 69.6% and a low false positive rate of 24.6% in predicting the questionnaire response. The mean t-score was 48.6 (*SD* 12.3) for the response 'No' and 60.9 (*SD* 14.4) for the response 'Yes'. These values were significantly different, $F(1,208) = 45.052$, $p < .001$.

Table 1 shows the agreement between the question 'Do you have trouble sleeping?' and the APS sleep disorder classification which was categorised as binary in order to compare agreement. The percentages are total percentages across the categories. The percent in agreement is moderate at 70.6% but with a kappa value of 0.39 indicating poor agreement because 20.6% of participants had a normal APS classification but reported trouble sleeping. There was a significant association between the items ($p < .001$).

Use of Alcohol: *Safe versus Hazardous/harmful Drinking*

Young offenders were questioned in detail about their alcohol consumption in the HQ and on the basis of their responses were classified into safe or hazardous/harmful drinking according to the 2001 National Health and Medical Research Council of Australia guidelines (National Health and Medical Research Council of Australia 2001). Of the 224 responses, 170 (75.9%) were classified as safe and 54 (24.1%) were classified as hazardous/harmful drinkers.

The agreement between the classification "Safe versus hazardous/harmful drinking" and the APS t-score was examined using a ROC curve. The APS t-score was a significant predictor of the "Safe/hazardous or harmful drinking". The area under the curve was moderate at 0.698 (95% CI 0.61, 0.78). The optimum cut-off t-score value for predicting harmful/hazardous drinking was 72.5, which gave a true positive rate of 59.3% and a low false positive rate of 24.7% in predicting the questionnaire response. The clear separation between mean t-scores between drinking levels is shown in Figure 1. The mean t-score was 62.2 (*SD* 18.1) for the safe group, and 75.9 (*SD* 24.6) for the hazardous/harmful group which was significantly different, $F(1,206) = 18.818$, $p < 001$.

Table 1 also shows the agreement between the "Safe versus hazardous/harmful drinking" classification and the APS substance abuse classification. The percent in agreement is 58.6% and the kappa value is poor at 0.23 reflecting the 37.0% of participants who were classified as sub-clinical to severe by the APS but who were classified as having a safe drinking level by the HQ. However, only 4.4% of participants were classified as normal by the APS but as having a hazardous/harmful drinking level by the HQ. There was a significant association between the classifications ($p < .001$).

Table 1

Percent Agreements for sleep, drinking and sexual and physical abuse

'Do you have trouble sleeping?' and the APS Sleep Disorder Classification	APS sleep disorder classification [n;%]		
Do you have trouble sleeping?	Normal	Sub-clinical to severe	Total
No	96	18	114
	47.1	8.8	55.9
Yes	42	48	90
	20.6	23.5	44.1
Total	138	66	204
	67.7	32.3	100.0
"Safe versus hazardous/harmful drinking" Classification and the APS Substance Abuse Classification	APS substance abuse classification		
Safe versus hazardous/harmful drinking	Normal	Sub-clinical to severe	Total
Safe	76	75	151
	37.4	37.0	74.4
Hazardous/harmful	9	43	52
	4.4	21.2	25.6
Total	85	118	203
	41.8	58.2	100.0

Sexual and Physical Abuse

Eighteen (8.2%) of the sample responded 'yes' to the question, 'Have you ever had sex against your will?' 65 (30.7%) responded 'yes' to the question, 'In the past 12 months have you had a physical injury that was deliberately caused by another person?' and 34 (15.2%) responded yes to the question, 'In the past 12 months, did any person affected by alcohol or drugs physically abuse you?' These questions were validated against the CTQ sexual abuse and CTQ physical abuse scales respectively.

The agreement between the sexual abuse question and CTQ sexual abuse scale and between each of the physical abuse questions and the CTQ physical abuse scale is shown in Table 2. There was a high percent in agreement at 91.3% and a moderate kappa value of 0.45. Kappa values are influenced by the distribution of responses and are rarely high in situations such as this where the prevalence of the outcome, in this case sexual abuse, is relatively low. There was a significant association between the two measurements ($p < .001$). The percent in agreement for the question "In the past 12 months have you had a physical injury that was deliberately caused by another person?" was 59.6% but the kappa value was poor at 0.12 reflecting the high percentages of participants who did not fall on the diagonal cells. There was no significant association between the two measurements ($p = .10$). The percent in agreement for the question "In the past 12 months, did any person affected by alcohol or drugs physically abuse you?" was 61.4% but again with a poor kappa value of 0.12 reflecting the high percentages of participants who did not fall on the diagonal cells. There was no significant association between the two measurements ($p = .057$).

Table 2

Percent Agreement between the Question about Sexual and Physical Abuse, and the CTQ Abuse Scales

HQ Question	HQ Response	CTQ Classification [n;%] None	CTQ Classification [n;%] Low-severe	Total
Have you ever had sex against your will?	No	180	13	193
		87.0	6.3	75.7
	Yes	5	9	14
		2.4	4.3	24.3
	Total	185	22	207
		94.0	6.0	100.0
In the past 12 months, have you had a physical injury that was deliberately caused by another person?	No	89	52	141
		45.0	26.3	71.3
	Yes	28	29	57
		14.1	14.6	28.7
	Total	117	81	198
		59.1	40.6	100.0
In the past 12 months, did any person affected by alcohol or drugs physically abuse you?	No	111	68	179
		52.8	32.4	85.2
	Yes	13	18	31
		6.2	8.6	14.8
	Total	124	86	210
		59.0	41.0	100.0

Note. Figures are the number in each group with the percent of the total sample below.

Anorexia and Bulimia

Table 3 shows the frequency of responses to the questions about anorexia and bulimia. Only four participants reported fasting and none reported vomiting or taking laxatives to lose weight or prevent weight gain.

Table 3

Frequency of Responses to Questions about Anorexia and Bulimia

Question	Response	Frequency N (%)
How do you describe your weight? (N=199)	Slightly/very underweight	44 (22.1)
	About the right weight	108 (54.3)
	Slightly/very overweight	47 (23.6)
Which of the following are you trying to do about your weight? (N=182)	Lose weight	14 (7.7)
	Stay the same	93 (51.1)
	Gain weight	75 (41.2)
In the last 4 weeks, did you eat less food, fewer calories or foods low in fat to lose weight or keep from gaining weight? (N=198)	No	178 (89.9)
	Yes	20 (10.1)
In the last 4 weeks, did you go without eating for 24 hours or more, also called fasting, to lose weight or to keep from gaining weight? (N=219)	No	215 (98.2)
	Yes	4 (1.8)
In the last 4 weeks, did you vomit or take laxatives to lose weight or to keep from gaining weight? (N=199)	No	199 (100.0)
	Yes	0 (0)

When the question of how participants described their weight was compared against the APS anorexia scale, only two participants (1%) who felt they were underweight were classified as sub-clinical-severe anorexic, seven (3.5%) who felt they were the right weight and seven (3.5%) who felt they were overweight. Similarly, for the question about how participants were managing their weight, only five (2.7%) who were trying to lose weight were classified as sub-clinical-severe anorexic and three (1.6%) who were trying to gain weight. Only four participants (2.0%) who were classified as sub-clinical-severe anorexic had eaten less food to lose weight in the last four weeks and none had fasted or vomited to keep from gaining weight.

There was high, and significant, agreement between BMI and how participants described their weight, $F(4,192) = 13.854$, $p < .001$, as shown in Figure 1 by an error bars plot. The difference in mean BMI between the lowest and the highest categories was 15.7 units. In addition, there was good agreement between BMI and how participants were trying to manage their weight. Participants who wanted to loose weight had a significantly higher BMI, $F(3,177) = 14.770$, $p < .001$, with a difference in BMI of 6.9 units between participants who wanted to lose or gain weight, and with the other two groups who wanted to maintain their weight having intermediate mean BMI values. There was also good agreement between mean BMI and trying to eat less food or calories. The mean BMI of the group who did not try to eat less food or calories was 23.2 (*SD* 4.0) compared to 29.6 (*SD* 5.7) in the group who did try to eat less food or calories. These group were significantly different, $F(1,193) = 34.823$, $p < .001$. There were insufficient responses for the questions about fasting or bulimia to validate the responses against BMI.

Anxiety and Self-Harm

Of the 218 responding to the question, ''Are you ever nervous?'' 126 (57.8%) replied ''None of the time''; 80 (36.7%) ''Some of the time''; and 12 (5.5%) ''Most/all of the time''. Of the 200 responding to the question, ''Have you ever

intentionally or deliberately hurt or injured yourself?" 174 (87.0%) replied 'no' and 26 (13.0%) replied 'yes'. In response to the question, "Have you ever seriously considered attempting suicide?" 160 (80.0%) replied 'no' and 40 (20.0%) replied 'yes'.

The percent in agreement for the question "Are you ever nervous?" when compared with the APS anxiety scale is shown in Table 4. The percent in agreement was 68.0% with a poor kappa value of 0.30 reflecting the 26.0% percent of participants who reported some anxiety but were normal on the APS anxiety scale. There was a significant association between the two measurements ($p < .001$).

The percent in agreement for the questions about intentional or deliberate self-harm or seriously considering suicide when compared with the APS suicide scale are also shown in Table 4. The percent in agreement was high at 89.0% for the self-harm question and 86.0% for the suicide questions. Both questions had a moderate kappa value of 0.50 and 0.49 respectively and both associations were significant ($p < .001$). A total of 6% of participants who answered 'Yes' to the self-harm question and 11% of participants who answered Yes to the suicide question were classified as normal on the APS suicide scale.

Table 4

Percent Agreement between "Are you ever nervous?" and the APS Anxiety Scale		APS anxiety scale [n ;%]		
Are you ever nervous?		Normal	Sub-clinical to moderate	Total
None of the time		103	12	115
		51.5	6.0	57.5
Some/most/all of the time		52	33	85
		26.0	16.5	42.5
Total		155	45	200
		77.5	22.5	100.0
Percent Agreement between Questions about Suicide and the APS Suicide Scale				
		APS Suicide Scale [n ;%]		
HQ Question	HQ Response	Normal	Sub-clinical to moderate	Total
Have you ever intentionally or deliberately hurt or injured yourself?	No	164	10	174
		82.0	5.0	87.0
	Yes	12	14	26
		6.0	7.0	13.0
	Total	176	24	200
		88.0	12.0	100.0
Have you ever seriously considered attempting suicide?	No	154	6	160
		77.0	3.0	80.0
	Yes	22	18	40
		11.0	9.0	20.0
	Total	176	24	200
		88.0	12.0	100.0

Note. Figures are the number in each group with the percent of the total sample below.

Discussion

The statistics obtained to validate questions used on the HQ are summarized for both categorical and continuous comparison variables in Table 5 and Figure 1. In general, the questions on trouble sleeping, hazardous drinking, sexual and physical abuse, perception of weight, anxiety and self-harm all have moderate to high percentages in agreement and significant associations between the HQ and the validation tools. Where ROC curves were used, the areas under the curve were also moderate to high indicating close associations. These results indicate that the HQ questions could all be used reliably. The questions with a percent in agreement less than 50% are not good markers of the instruments against which they were validated.

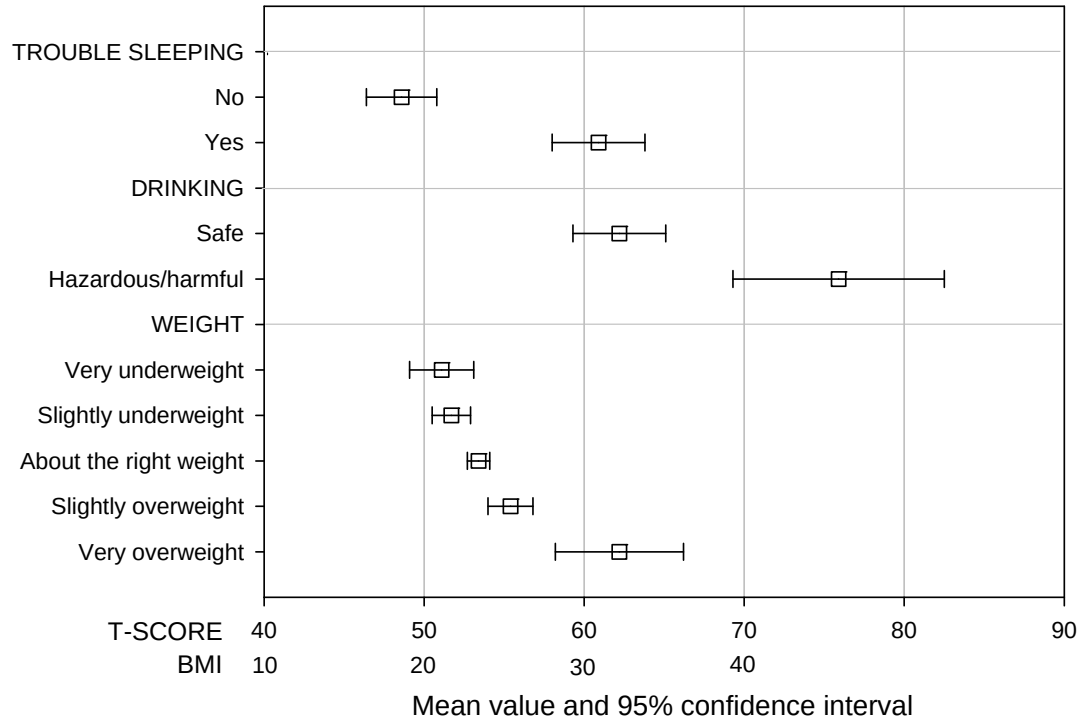
Table 5

Summary of Validation Statistics

Question	Categorical			Continuous	
	Percent in agreement	Kappa	*p* value	Area under curve	*p* value
Do you have trouble sleeping? (versus APS)	70.6	0.39	<.001*	0.78	<.001*
Safe versus hazardous/ harmful drinking (versus APS)	58.6	0.23	<.001*	0.698	<.001*
Have you ever had sex against your will? (versus CTQ)	91.3	0.45	<.001*	_	_
In the past 12 months, have you had a physical injury deliberately caused by another person? (versus CTQ)	59.6	0.12	.10	_	_
In the past 12 months, did any person … physically abuse you? (versus CTQ)	61.4	0.12	.057	_	_
How do you describe your weight? (versus BMI)	_	_	_	_	<.001*
Which of the following are you trying to do about your weight? (versus BMI)	_	_	_	_	<.001*
In the last 4 weeks, did you eat less food to lose weight or keep from gaining weight? (versus BMI)	_	_	_	_	<.001*
Are you ever nervous? (versus APS)	68.0	0.30	<.001*	_	_
Have you ever intentionally or deliberately hurt or injured yourself? (versus APS)	89.0	0.50	<.001*	_	_
Have you ever seriously considered attempting suicide? (versus APS)	86.0	0.49	<.001*	_	_

Note. Dashes indicate that the statistic was not able to be calculated.
*$p < .05$.

Figure 1 Error bars plots showing separation in mean t-scores for sleeping, drinking and weight items between the questionnaire levels.



Adolescent offenders were accurate in their self-report of their physical and mental health, comparable with community and clinical samples of adolescents (Lipschitz et al. 1999; Decaluwe et al. 2004). Lower accuracy or agreement rates between self-reports and interviews of suicidality or self-harm in adolescent clinical and outpatient samples (Velting et al. 1998) have been attributed to problematic definitions and difficulties distinguishing between suicidal behaviors and other behaviors associated with intense emotional distress. Adolescent offenders in this sample were accurate disclosers of current suicidal ideation and sexual abuse because they were offered clear definitions and the opportunity to disclose under conditions of anonymity (Safer 1997), using valid measures for comparison (Bernstein, et al. 1997).

The majority of comparisons between factors using kappa statistics indicated moderate levels of agreement. Low agreement and non significant associations for the physical abuse domain are more likely to due to a discrepancy between question content and form rather than inaccurate self-reports. Studies of the reliability of reported health behaviors from adolescent community and clinical samples showed that a lack of inter-changeability between questions used for comparison resulted in underestimations of agreement and associations between measures (Field et al. 2004; Shillington et al. 2000). Further, factors such as physical abuse demonstrated lower agreement as kappa is influenced by the distribution of responses and is rarely high when the expected prevalence of the outcome, such as abuse, is low (Altman 1996).

The majority of results indicated only moderate levels of agreement using the kappa Statistic. However, kappa only increases as the proportion of negative and positive responses become more equal for each self-report item. As unequal proportions of response were evident in the HQ, both kappa and the proportion in agreement are used to jointly assess which questions were most reliable. False positive results, although small, were observed for the "trouble sleeping" and the

"safe versus hazardous/harmful drinking" items when compared to the respective APS classifications. However, when examining the accuracy of self-reports obtained from measures used only to screen for behavior, the true positive rate is paramount and less emphasis is placed on slight overestimations of behavior prevalence (Field et al. 2004).

Methodological advantages for exploring the accuracy of adolescent self-report in this study included a large sample size for most comparisons, a method of data collection that stressed confidentiality and anonymity (Johnson et al. 2004; Lintonen et al. 2004; Decaluwe et al. 2004; Tanofsky-Kraff, Morgan et al. 2003), and administration by experienced and well-trained interviewers using a manualized protocol to ensure consistency in questionnaire delivery. The study was not specifically designed to assess accuracy of self-report and respondents were not aware that comparisons would be made of their self-report consistency across different forms of assessment. Awareness that such comparisons may be made has been shown to distort the veracity of responses by providing an incentive to report honestly (Del Boca et al. 2003; Johnson et al. 2000). Other methodological strengths of the study were outlined in the methods.

Limitations

The retrospective study design did not permit analysis of possible mediating variables or individual differences (eg cognitive capacity, education, cultural factors) known to influence participants' responding (Del Boca et al. 2003; Lintonen et al. 2004; Passi et al. 2003; Velting et al. 1998). Likewise, the small number of females in the sample did not permit an analysis of the potential influence of gender on responding. Future research should explore self-reports obtained from other juvenile offender samples (e.g. community-based juvenile offenders) to investigate the generalizability of these findings. Further exploration of the accuracy of juvenile offender self-reports in other domains of interest and using other methods of corroboration would strengthen conclusions about accuracy of self-report in this population. Independent evidence is often difficult to obtain and there is currently a lack of consensus regarding management of discrepancies across measures (Del Boca et al. 2003; Johnson et al. 2004; Lintonen et al. 2004; Williams et al. 2005; Decaluwe et al. 2004; Ferdinand et al. 2004; Tanofsky-Kraff, Morgan et al. 2003). The instruments used in this study have robust psychometric properties including systematic response bias detection scales and were therefore appropriate standards to use for comparison of self-report to face-to-face interview questions. The use of all available data and multiple sources of direct and objective information, along with analyses that are sensitive to changes in agreement, such as kappa used in the current study, will strengthen the implications of the present results (Johnson et al. 2004; Williams et al. 2005; Winegar et al. 1999; Percy et al. 2005).

Conclusions

The findings from this study are consistent with the majority of previous research on adolescent populations showing moderate to high reliability in their self-report of a variety of health behaviors. The self-report questions used in this study could all be used reliably and support the continued use of self-report as a data collection method in routine assessments and practice evaluations and as the basis for clinical formulation and treatment planning. Future research could strengthen these conclusions by implementing alternate research designs, employing multiple, objective and completely interchangeable measures on which to base comparisons, and examining a wider range of health behaviors.

References

1. Altman, D.G. (1996). *Inter-rater agreement in practical statistics for medical research*. London: Chapman and Hall.
2. Altman, D.G. & Bland, J.M. (1994). Diagnostic tests 3: Receiver operating characteristics plots. *British Medical Journal,* 309, 188.
3. Bernstein, D. & Fink, L. (Eds) (1998). *Manual for the childhood trauma questionnaire*. New York: The Psychological Corporation.
4. Bernstein, D.P., Ahluvalia, T., Pogge, D. & Handlesman, L. (1997). Validity of the childhood trauma questionnaire in an adolescent psychiatric population. *Journal of the American Academy of Child and Adolescent Psychiatry,* 36, 340-348.
5. Carroll, K.M. (1995). Methodological issues and problems in the assessment of substance use. *Psychological Assessment,* 7, 349-358.
6. Decaluwe, V. & Braet, C. (2004). Assessment of eating disorder psychopathology in obese children and adolescents: Interview versus self-report questionnaire. *Behavior Research and Therapy*, 42, 799-811.
7. Del Boca, F.K. & Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction,* 98, 1-12.
8. Fairburn, C.G. & Beglin, S.J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders,* 16, 363-370.
9. Fendrich, M. & Xe, Y. (1994). Validity of drug use reports from juvenile arrestees. *International Journal of the Addiction*, 29, 971-985.
10. Ferdinand, R.F., Ende, J.V.D. & Verhulst, F.C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology,* 113, 198-206.
11. Field, A.E., Taylor, C.B., Celio, A. & Colditz, G.A. (2004). Comparison of self-report to interview assessment of bulimic behaviors among preadolescent and adolescent girls and boys. *International Journal of Eating Disorders,* 35, 86-92.
12. Gadow, K.D., Sprafkin, J., Carlson, G.A., Schneider, J., Nolan, E.E., Mattison, R.E., et al. (2002). A DSM-IV-referenced, adolescent self-report scale. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 671-679.
13. Johnson, M.E., Fisher, D.G., Montoya, I., Booth, R., Rhodes, F., Andersen, M., et al. (2000). Reliability and validity of not-in-treatment drug users' follow-up self-reports. *AIDS & Behavior*, 4, 373-380.
14. Johnson, P.B. & Richter, L. (2004). Research note: What if we were wrong? Some possible implications of systematic distortions in adolescents' self-reports of sensitive behaviors. *Journal of Drug Issues,* 34, 951-970.
15. Langhinrichsen-Rohling, J., Arata, C., Bowers, D., O'Brien, N. & Morgan, A. (2004). Suicidal behavior, negative affect, gender, and self-reported delinquency in college students. *Suicide and Life-Threatening Behavior*, 34, 255-266.
16. Lintonen, T., Ahlstrom, S. & Metso, L. (2004). The reliability of self-reported drinking in adolescence. *Alcohol & Alcoholism*, 39, 362-368.
17. Lipschitz, D.S., Bernstein, D.P., Winegar, R.K. & Southwick, S.M. (1999). Hospitalized adolescents' reports of sexual and physical abuse: A comparison of two self-report measures. *Journal of Traumatic Stress,* 12, 641-653.
18. National Health and Medical Research Council of Australia. (2001). Australian Alcohol Guidelines: *Health risks and benefits* 2001. Canberra, Australia: Author.
19. NSW Department of Juvenile Justice (2003). NSW *Young people in custody health survey - key findings report* (ISBN: 0 7347 6518 5). Sydney: Author.
20. Ohida, T., Osaki, Y., Doi, Y., Tanihata, T., Minowa, M., Suzuki, K., et al. (2004). An epidemiologic study of self-reported sleep problems among Japanese adolescents. *Sleep*, 27(5), 978-985.
21. Paivo, S.C. (2001). Stability of retrospective self-reports of child abuse and neglect before and after therapy for child abuse issues. *Child Abuse & Neglect*, 25, 1053-1068.
22. Passi, V.A, Bryson, S.W. & Lock, J. (2003). Assessment of eating disorders on adolescents with anorexia nervosa: Self-report questionnaire versus interview. *International Journal of Eating Disorders,* 33, 45-54.

23. Validity. (2001). In Peat, J.K., Mellis, C.M., Williams, K. & Xuan, W (Eds.), *Health science research. A handbook of quantitative methods* (pp. 105-113). Sydney, Australia: Allen and Unwin.
24. Percy, A., Mcalister, S., Higgins, K., McCrystal, P. & Thornton, M. (2005). Response consistency in young adolescents' drug use self-reports: A recanting rate analysis. *Addiction, 100*, 189-196.
25. Reynolds, W.M. (1998). *Adolescent psychopathology scale: Administration and interpretation manual*. U.S.A: Psychological Assessment Resources Inc.
26. Safer, D.J. (1997). Self-reported suicide attempts by adolescents. *Journal of Clinical Psychiatry*, 9, 263-269.
27. Santelli, J.S., Klein, J.D., Graff, C.A., Allen, M.J. & Elster, A.B. (2002). Reliability in adolescent reporting of clinical counseling, health care use and health behaviors. *Medical Care*, 40, 26-37.
28. Shillington, A.M. & Clapp, J.D. (2000). Self-report stability of adolescent substance use: are there differences for gender, ethnicity and age? *Drug and Alcohol Dependence*, 60, 19-27.
29. Statistical Package for the Social Sciences (Version 12 for windows) (2003) [Computer software]. Illinois: SPSS Inc.
30. Strand, V.C, Sarmiento, T.L. & Pasquale, L.E. (2005). Assessment and screening tools for trauma in children and adolescents: a review. *Trauma, Violence and Abuse,* 6, 55-78.
31. Tanofsky-Kraff, M., Morgan, C.M., Yanovski, S.Z., Marmarosh, C., Wilfley, D.E. & Yanovski, J.A. (2003). Comparison of assessments of children's eating-disordered behaviors by interview and questionnaire. *International Journal of Eating Disorders*, 33, 213-224.
32. Tanofsky-Kraff, M., Yanovski, S.Z. & Yanovski, J.A. (2005). Comparison of child interview and parent reports of children's eating disordered behaviors. *Eating Behaviors,* 6, 95-99.
33. Velting, D.M., Rathus, J.H. & Ansis, G.M. (1998). Asking adolescents to explain discrepancies in self-reported suicidality. *Suicide and Life-Threatening Behavior*, 28, 187-196.
34. Waters, E., Salmon, L.A., Wake, M., Wright, M. & Hesketh, K.D. (2001). The health and well-being of adolescents: a school-based population study of the self-report Child Health Questionnaire. *Journal of Adolescent Health*, 29, 140-149.
35. Waters, E., Stewart-Brown, S. & Fitzpatrick, R. (2003). Agreement between adolescent self-report and parent reports of health and well-being: Results of an epidemiological study. *Child Care, Health & Development,* 29, 501-509.
36. Wekerle, C., Wolfe, D.A., Hawkins, D.L., Pittman, A.L., Glickman, A. & Lovald, B.E. (2001). Childhood maltreatment, posttraumatic stress symptomatology, and adolescent dating violence: Considering the value of adolescent perceptions of abuse and a trauma mediation model. *Development and Psychopathology,* 13, 847-871.
37. Williams, R.J. & Nowatzki, N. (2005). Validity of adolescent self-report of substance use. *Substance Use and Misuse,* 40, 299-311.
38. Winegar, R.K. & Lipschitz, D.S. (1999). Agreement between hospitalized adolescents' self-reports of maltreatment and witnessed home violence and clinician reports and medical records. *Comprehensive Psychiatry*, 40, 347-352.
39. Wolfson, A.R., Carskadon, M.A., Acebo, C., Seifer, R., Fallone, G., Labyak, S.E., et al. (2003). Evidence for the validity of a sleep habits survey for adolescents. *Sleep*, 26, 213-216.
40. Yacoubian, G.S., Vanderwall, K.L. & Johnson, R.J., Urbach, B.J. & Peters, R. (2003). Comparing the validity of self-reported recent drug use between adult and juvenile arrestees. *Journal of Psychoactive Drugs*, 35, 279-284.