

Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment

A. Alyahri¹ and R. Goodman¹

التحقق من صحة النسخة العربية للاستبيان الخاص بقياس أوجه القوة والمصاعب، وتقييم حالة النماء والمعاودة

عبد الله أحمد اليهري، روبرت قدمان

الخلاصة: قام الباحثان باختبار صحة النسخ العربية للمقياسين الأساسيين المستخدمَين في قياس الباثولوجيا النفسية للطفل، وهما: الاستبيان الخاص بقياس أوجه القوة والمصاعب (SDQ)، وتقييم حالة النماء والمعاودة (DAWBA). وقد تم توزيعهما على أولياء أمور ومعلمي عيّنَين من الأطفال اليمنيين، تتراوح أعمارهم بين 5 و12 عاماً، إحداهما من عيادات نفسانية (العدد 87) والأخرى من المجتمع بصفة عامة (العدد: 100). وقد تمايزت درجات الاستبيان الخاص بقياس أوجه القوة والصعوبات (SDB) تمايزاً كبيراً بين العيّنتين، وكذلك بين الأطفال المشخصين بحالات نفسانية متفاوتة. وقد أظهر تقييم حالة النماء والمعاودة (DAWBA) اتفاقاً كبيراً مع تشخيص العيادات المستقلة. وعلى هذا فإن الإيجاز الذي يتسم به استبيان قياس أوجه القوة والمصاعب، واللقاءات التي أُجريت لتقييم حالة النماء والمعاودة المعتمدة بطبيعتها على ردود المستجيب، تجعل هاتين الأدواتِ مناسبتين تماماً للاستخدام في البلدان التي تعاني من نقص شديد في القوى البشرية المدربة.

ABSTRACT We examined the validity of the Arabic versions of 2 main measures of child psychopathology: the Strengths and Difficulties Questionnaire (SDQ) and the Development and Well-Being Assessment (DAWBA). They were administered to the parents and teachers of 2 samples of 5–12-year-old Yemeni children, one from psychiatric clinics ($n = 87$) and the other from the community ($n = 100$). The SDQ scores distinguished well between the 2 samples and also between children with different psychiatric diagnoses. The DAWBA showed substantial agreement with independent clinic diagnosis. The brevity of the SDQ and the respondent-based nature of the DAWBA interview make these tools feasible for use in countries where there is a severe shortage of skilled manpower.

Validation de la version arabe des questionnaires *Strengths and Difficulties Questionnaire* et *Development and Well-Being Assessment*

RÉSUMÉ Nous avons examiné la validité de la version arabe de 2 instruments de mesure de la psychopathologie infantile : *Strengths and Difficulties Questionnaire* (SDQ) et *Development and Well-Being Assessment* (DAWBA). Ces instruments ont été utilisés avec les parents et les enseignants de 2 échantillons d'enfants yéménites âgés de 5 à 12 ans, l'un issu des consultations psychiatriques ($n = 87$) et l'autre de la communauté ($n = 100$). Les scores au SDQ établissaient une nette distinction entre les 2 échantillons et aussi entre les enfants ayant des diagnostics psychiatriques différents. Le DAWBA montrait une grande concordance avec le diagnostic clinique indépendant. Le SDQ est un questionnaire bref et l'entretien DAWBA est de type structuré, ce qui rend ces instruments utilisables dans les pays où il y a une grave pénurie de personnel qualifié.

¹Department of Child and Adolescent Psychiatry, Institute of Psychiatry, King's College, London, United Kingdom (Correspondence to A. Alyahri: alyahri32hotmail.com).

Introduction

There is growing recognition of the importance of child psychiatric disorders in developing countries. These disorders are important not only because they result in suffering for children and those around them, but also because they interfere with social and educational development, and can lead to life-long social and psychiatric problems [1]. There is a pressing need, particularly in developing countries with very limited access to child mental health professionals, to develop simple screening mechanisms to help ensure that referrals to child mental health services are appropriate. It would be unrealistic to develop screening mechanisms that depend on complex or expensive measures and have to be administered by highly trained staff [2].

The Strengths and Difficulties Questionnaire (SDQ) is a brief tool that could be a relatively cheap and easy screening measure. It was originally published in English [3], and has subsequently been translated into over 60 languages. The fact that the SDQ is predictive of psychiatric diagnoses in many developed and developing countries [4] raises the possibility that it might be useful as a screen for psychiatric disorders in community settings, primary health care or paediatric clinics in Yemen.

The Development and Well-Being Assessment (DAWBA) is a package of questionnaires and structured interviews that also collects answers to open-ended questions for clinical review. In developing countries, where there is a severe shortage of skilled manpower, use of structured interviews is the most cost-effective and feasible way of carrying out community surveys [5]. The DAWBA was initially designed for a nationwide epidemiological survey of common emotional and behavioural disorders in Britain [6]. Previous studies have provided

evidence for the validity of the DAWBA in a developed country [6] as well as in a developing country [7].

Translating established measures is generally quicker and cheaper than developing new measures for each new language or country, and international measures have the additional advantage of facilitating international comparison. It is important, however, that standardized measures developed outside a particular culture and language community should be validated before their use in that setting [5].

This study aimed to validate the Arabic versions of standardized child psychiatric measures (the SDQ and the DAWBA) for future clinical and research application in Yemen and elsewhere in Arabic speaking countries.

Method

Clinical sample

The clinical sample was obtained from Aden Neuropsychiatric Teaching Hospital, the main psychiatric hospital in Yemen, Alwahda Paediatric Teaching Hospital, the biggest paediatric hospital in Yemen, and from school-based psychiatric clinics in Aden. These hospitals receive referrals from a wide catchment area, mainly from Aden, Abian and Lahj provinces. The children referred to these clinics are of mixed socioeconomic status. The SDQ and the DAWBA were administered to parents. The interviewers were blind to the diagnosis made by the psychiatrist, the paediatrician or the psychologist. The SDQ was read out by the interviewer when the respondents' literacy skills were insufficient for them to complete the questionnaires directly. The SDQ and the DAWBA were administered to teachers as well. These measures were administered on a consecutive series of 108

patients aged 5–12 years when first seen at the clinics between February and July 2002. Eleven were excluded because missing answers made it impossible to generate all scores. The final sample consisted of the remaining 97 children. Teacher SDQs and DAWBAs were available only for 68 (70%), either because the child was not at school or because the parents did not consent to contact with the school.

The mean age of the children in the sample was 9.4 (standard deviation 3.1) years and 53 (54.6%) were male.

Community sample

The community sample was selected from schools in 2 different areas, Crater and Sheikh Othman, of Aden city. These areas were chosen to represent families whose socioeconomic status was similar in range to those who made up the clinic sample. One hundred children between 5 and 12 years were selected through a 2-stage sampling programme. In the first stage, a clustered random sample of classes from the first 4 grades of primary school was selected from each school and in the second stage, children were randomly selected, 3 or 4 from each class, from the class registers of 24 classes in 3 boys' schools and 3 girls' schools; the rest were selected from 1 kindergarten.

Parents were visited at home or seen in schools and all agreed to take part in the study. Parents and teachers completed SDQs, with the questionnaire being read out when the respondent did not have adequate literacy skills. Complete parent and teacher SDQ information was available on all 100 children. The mean age of the sample was 8.9 (standard deviation 1.6) years and 51% were males. The community and clinic samples were matched for sex ($\chi^2 = 0.63$, $df = 1$, $P = 0.4$) and did not differ significantly

for age ($P = 0.09$; 95% confidence interval -0.99 – 0.07)

Assessments were carried out between January 2002 and April 2002, avoiding the first term when teachers do not yet know their pupils, and avoiding the end of the last term when teachers are often busy with the end-of-year examinations.

Inclusion/exclusion criteria

Clinical sample

Children between the age of 5 and 12 years who had been referred to the psychiatric clinic were included. All consecutive referrals were eligible except for children with moderate or severe learning disability or whose only problem was enuresis, epilepsy or a specific learning disorder.

Community sample

All children between 5 and 12 years of age attending the kindergarten and the first 4 grades of the 7 selected schools were eligible.

Instruments

The SDQ is a brief behavioural screening questionnaire that covers 25 attributes, some positive and others negative [3]. The 25 items are divided between 5 scales of 5 items each, generating scores for conduct problems, inattention-hyperactivity, emotional symptoms, peer problems and prosocial behaviour. All scales but the last are summed to generate a total difficulties score (range 0–40). The same questionnaire can be completed by the parents or the teachers of 4–16-year-olds [3]. Besides covering common areas of emotional and behavioural difficulties, it also inquires whether the informant thinks that the child has a problem in these areas and, if so, asks about resultant distress and social impairment [8]. The web site at <http://www.sdqinfo.com>

provides more information and downloadable questionnaires in many languages and scoring instructions.

The SDQ had previously been translated into Arabic. This translation was revised to maximize understanding in Yemen, and back-translated to ensure fidelity to the original English version.

The DAWBA is a package of questionnaires, interviews and rating techniques designed to generate ICD-10 and DSM-IV psychiatric diagnoses on 5–16-year-olds [9,10]. It is designed so that non-clinical interviewers can administer a structured interview to parents about psychiatric symptoms and resultant impact. When definite symptoms are identified by the structured questions, interviewers use open-ended questions and supplementary prompts to get parents to describe the problems in their own words. These descriptions are transcribed verbatim by the interviewers but are not rated by them. Teachers complete a brief questionnaire covering the main conduct, emotional, and hyperactivity symptoms and any resultant impairment. The different sorts of information are brought together by a computer program that also predicts likely diagnoses. These computer-generated summary sheets and diagnoses form a convenient starting point for experienced clinical evaluators, who decide whether to accept or overturn the computer diagnosis (or lack of diagnosis) in the light of their review of all the data, including transcripts [6].

The DAWBA was translated from English to Arabic by the first author. To ensure translation equivalence, a back-translation was done in the Faculty of Linguistics and Translation, Ajman University, United Arab Emirates. Conceptual and linguistic problems were resolved by extensive consultation with the second author, who wrote the original version, and a number of Arab psychiatrists. The translated version was pi-

lotted on a number of mothers from various Arab countries at the Islamic Welfare Centre in London. The web site at <http://www.dawba.com> provides more information, and downloadable versions of the interviews and questionnaires in many languages.

Selection and training of interviewers

Four female interviewers (participation rate was expected to be higher with females) were selected, 3 were experienced clinical psychologists and 1 was a psychiatric senior house officer. The mode of introducing the DAWBA and the SDQ was explained in detail. Training on the theoretical aspects was carried out for 1 week then role-play exercises were carried out followed by a field-training session in the clinic.

Clinical diagnosis

Children from the psychiatric and the paediatric clinics were assigned clinical diagnoses based on the operationalized criteria of either ICD-10 or DSM-IV. These clinical diagnoses were made at the time of initial assessment by a psychiatrist, a clinical psychologist or a paediatrician. The diagnoses were made blind to the children's SDQ and DAWBA assessments.

Diagnoses were collapsed into 2 broad categories to provide cell sizes that would be sufficient for meaningful analysis and also to avoid misclassification of subcategory diagnoses, as the initial assessments were carried out by a variety of professionals at different levels of seniority. The categories were *externalizing disorder* (including hyperkinetic, conduct disorder and oppositional disorder) and *emotional disorder* (including anxiety, depressive disorder and obsessive compulsive disorder). Overall, 30 children had an emotional disorder and 79 had an externalizing disorder (with 12 individuals having both).

Development and Well-Being Assessment diagnosis

All open-ended comments were translated into English by one of the authors. Cultural and linguistic nuances in each case were discussed by the authors, after which the DAWBA diagnosis was made by the second author, who had previously made or supervised many thousands of DAWBA diagnoses. The rating consisted of reviewing information from the structured and open-ended questions and re-evaluating the computer diagnosis to provide a final clinical diagnosis based on DSM-IV and ICD-10 diagnostic criteria for each subject [9,10]. Diagnoses were collapsed into 2 broad categories to match the diagnostic categories adopted for the clinical diagnosis, sufficient for meaningful analysis.

Statistical analysis

Validity of the Development and Well-Being Assessment

Based on criteria originally proposed by Landis and Koch [11], the validity of the translated version was primarily tested by examining the correspondence between clinical diagnosis and DAWBA diagnosis, using the kappa coefficient (κ , external validity). If there is complete agreement, then $\kappa = 1$. If there is no more agreement than would be expected by chance alone, then $\kappa = 0$.

Validity of the Strengths and Difficulties Questionnaire

The ability of different SDQ scales to distinguish between community and clinic subjects was examined using receiver operating characteristic (ROC) curves, employing the area under the curve as the index of discriminant ability. For this purpose, the underlying assumption was that the children in the clinic sample were substantially more likely to have psychiatric disorders than

were the children in the community sample (i.e. the relevant psychiatric disorders were more common in the high-risk than in the low-risk group). In ROC analyses, sensitivity (percentage of correctly identified "cases") and specificity (percentage of correctly classified healthy "non-cases") are calculated for all possible cut-off points of a score, and then combined in a single value called "area under the curve" (AUC). The AUC value obtained in this way reflects the discriminant validity. As a guide to interpretation, the AUC is 1.0 for a measure that discriminates perfectly and 0.5 for a measure that has no better than chance accuracy. With the number of subjects in this study, the level of significance is significantly better than chance when the AUC is ≥ 0.6 .

To generate ROC curves for each SDQ scale, the community sample was compared with the most relevant operationalized categorical diagnoses derived from the DAWBA. For 4 of the scales, the total difficulties scale, the total impact scale, the peer problems scale and the prosocial behaviour scale, the comparison was between all those in the community group and all those in the clinic group. The remaining 3 scales, emotional, conduct and hyperactivity symptoms, were judged by comparing the entire community sample with those clinic cases who had the corresponding disorder, as diagnosed by the DAWBA. For example, the discriminant power of the SDQ emotional scale was judged by comparing all community subjects with those children who had been diagnosed by the DAWBA as having an emotional disorder.

Results

Validity of the Arabic Development and Well-Being Assessment

DAWBA diagnosis was done on 86 (89%) of the 97 clinical cases. Table 1 shows the

Table 1 Development and Well-Being Assessment (DAWBA) and clinic diagnoses for emotional disorders

Clinic diagnosis of an emotional disorder	DAWBA diagnosis of an emotional disorder		Total
	No	Yes	
No	56	11	67
Yes	5	25	30
Total	61	36	97

$\kappa = 0.634$ (95% CI 0.473–0.795).

cross-tabulation of DAWBA and clinic diagnoses for emotional disorders. The DAWBA diagnosed the same disorder in 83% of cases.

Table 2 shows the cross-tabulation of DAWBA and clinic diagnoses for externalizing disorders (conduct/hyperkinesis). The DAWBA agreed with clinic diagnoses in 86% of cases. Overall there was substantial agreement between the DAWBA and the clinic diagnoses on the 2 main diagnostic groupings: emotional disorders and externalizing disorders.

Validity of the Arabic Strengths and Difficulties Questionnaire

Table 3 summarizes the ability of different SDQ scales and informants (parents and teachers) to distinguish between commu-

nity and clinic subjects, as gauged by the AUC. All 7 SDQ scales (total impact, total difficulties, emotional symptoms, conduct problems, hyperactivity, peer problems and prosocial behaviour) seem potentially useful for predictive purposes. In each case, the AUC was significantly greater than 0.5 ($P < 0.001$). Discrimination was best for hyperactivity and conduct scales for both the parent SDQ (AUC = 0.97 for hyperactivity and 0.88 for conduct scale) and the teacher SDQ (AUC = 0.97 for hyperactivity and 0.86 for conduct scale).

In the clinic sample, discrimination for emotional, conduct and hyperactivity scores between patients with different sorts of disorders (as defined by the DAWBA) was also examined using the AUC (Table 4). For example, the SDQ hyperactivity score discriminated well between patients with hyperkinetic disorder and psychiatric controls, i.e. clinic patients without a hyperkinetic disorder but with other diagnoses instead. Similarly, conduct and emotional scores all discriminated satisfactorily between clinic cases with and without the corresponding type of disorders. All AUCs represented a level of prediction substantially better than chance ($P < 0.001$).

It is clear that both the parent SDQ and the teacher SDQ were as good at discriminating between different types of disorder within the clinic sample as they were at distinguishing between the clinic and the community sample.

Discussion

The main purpose of this study was to investigate the validity of the Arabic versions of 2 main measures of child psychopathology, namely the SDQ and the DAWBA.

The DAWBA worked well judged by its agreement with independent clinical

Table 2 Development and Well-Being Assessment (DAWBA) and clinic diagnoses for externalizing disorders

Clinical diagnosis of an externalizing disorder	DAWBA diagnosis of an externalizing disorder		Total
	No	Yes	
No	29	5	34
Yes	9	54	63
Total	38	59	97

$\kappa = 0.691$ (95% confidence interval 0.542–0.840)

Table 3 Ability of different Strength and Difficulties Questionnaire (SDQ) scores to distinguish between community and clinic samples

SDQ score	Comparing community and clinic ^a samples			
	Parent rated		Teacher rated	
	AUC	95% CI	AUC	95% CI
Total impact	0.84 ^b	0.78–0.90	0.85 ^b	0.78–0.92
Total difficulties	0.81 ^b	0.74–0.87	0.76 ^b	0.68–0.83
Emotional symptoms	0.78 ^c	0.70–0.85	0.70 ^c	0.59–0.81
Conduct problems	0.88 ^d	0.82–0.93	0.86 ^d	0.79–0.92
Hyperactivity	0.97 ^e	0.93–1.00	0.97 ^e	0.94–1.00
Peer problems	0.70 ^b	0.62–0.77	0.66 ^b	0.57–0.74
Prosocial behaviour	0.78 ^b	0.71–0.85	0.65 ^b	0.56–0.74

CI = confidence interval.

^aThe clinic sample included all clinic cases for total impact and difficulties, peer problems and prosocial behaviour. For the remaining scores, only clinic cases with the corresponding diagnosis were included; e.g. the area under the curve (AUC) for emotional symptoms was for all community subjects compared with just those clinic cases who had an emotional disorder.

^bFor all 100 community participants versus all 86 clinic patients with any Development and Well-Being Assessment (DAWBA) diagnosis.

^cFor all 100 community participants versus all 36 clinic patients with DAWBA emotional diagnosis.

^dFor all 100 community participants versus all 57 clinic patients with DAWBA conduct diagnosis.

diagnosis. This study provided the first evidence for the validity of the Arabic version of the DAWBA. Furthermore, these results support findings from previous studies from other developing countries on the validity

of DAWBA parent interview. For example, in a Brazilian report, DAWBA made a diagnosis on 94% of a clinical sample who had an independent clinical diagnosis, with agreement on diagnostic grouping for

Table 4 Ability of different Strength and Difficulties Questionnaire (SDQ) scores to distinguish between disorders within the clinic sample

SDQ score	Comparing clinic cases	Parent rated		Teacher rated	
		AUC	95% CI	AUC	95% CI
Emotional symptoms	Emotional disorder (yes $n = 36$; no $n = 50$) ^a	0.76	0.66–0.86	0.72	0.60–0.85
Conduct problems	Conduct disorder (yes $n = 57$; no $n = 29$) ^b	0.89	0.83–0.96	0.89	0.82–0.97
Hyperactivity	Hyperkinesia (yes $n = 27$; no $n = 59$) ^c	0.86	0.78–0.95	0.89	0.81–0.97

AUC = area under the curve.

CI = confidence interval.

78% [12]. In a study in Bangladesh, there was substantial agreement between the DAWBA and the independent clinic diagnosis (κ 0.63–0.94) [13].

Our study examined validity rather than reliability. Of course, the evidence for validity provides indirect evidence for reliability too; an unreliable set of measures would also have done poorly on tests of validity [6].

The interviewers were encouraged to express their opinion and the families' feelings and attitudes about this relatively long interview as this was the first time a structured psychiatric interview had been used in Yemen and because of the intention to use this measure subsequently in a large-scale epidemiological study. In general, interviewers reported very encouraging attitudes from the families, and the interviewers found it an enjoyable experience for themselves as well because they felt they had acquired more knowledge about child psychiatric problems. The skip rules also made the interview shorter and easy to administer.

Using the SDQ, it was possible to discriminate well, with a fair degree of precision, between community subjects and clinic patients on the basis of all 7 SDQ scales. Within the clinic sample as well, the SDQs were able to predict broad-band psychiatric diagnoses with a fair degree of accuracy. This level of accuracy could potentially be clinically useful. For example, children whose parent and teacher SDQ scores suggest that they are at a particularly high risk of a hyperkinetic disorder could be allocated to a professional with particular expertise in this domain.

These results support findings from 2 previous studies conducted on the SDQ in

Arabic countries. The first was conducted in the Gaza strip on children in 4 age bands [14]. In spite of the small sample size in each age group, the findings indicated that the SDQ was very promising as a screening measure or rating scale. More recently, in Yemen, Almaqrami and Shuwail found that the self-report version of the SDQ discriminated appropriately between a clinic and a community sample and was capable of detecting childhood emotional and behavioural disorders in clinical settings [15].

Conclusion

The present study suggests that the Arabic version of the SDQ may predict psychiatric diagnosis accurately enough to be of value for screening and epidemiological studies as well as for clinical assessment, and shows that the SDQ is not only a practical and economical, but also valid measure for assessing different behavioural aspects of children.

The DAWBA shows a substantial agreement with independent clinical diagnoses. Such measures are essential to assess child psychiatric diagnoses in large-scale surveys to determine the prevalence of mental disorders in children in Yemen or other Arabic countries. The DAWBA may also be useful for standardized assessment within child mental health services.

Acknowledgements

This study was partially supported by the Eastern Mediterranean Regional Office of the World Health Organization. The authors are grateful to all parents, teachers and interviewers involved.

References

1. Rutter M. Connections between child and adult psychopathology. *European child and adolescent psychiatry*, 1996, 5(suppl. 1):4–7.
2. Mullick M, Goodman R. Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. *Social psychiatry and psychiatric epidemiology*, 2001, 36(2):94–9.
3. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, 1997, 38(5):581–6.
4. Woerner W et al. The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *European child and adolescent psychiatry*, 2004, 13(suppl. 2):47–54.
5. Wickramasinghe S et al. The Clinical Interview Schedule—Sinhala version: validation in a community setting in Sri Lanka. *International journal of methods in psychiatric research*, 2002, 11(4):169–77.
6. Goodman R et al. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of child psychology and psychiatry*, 2000, 41(5):645–55.
7. Fleitlich-Bilyk B, Goodman R. The prevalence of child and adolescent psychiatric disorders in southeast Brazil. *Journal of American Academy of Child and Adolescent Psychiatry*, 2004, 43(6):727–34.
8. Goodman R et al. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British journal of psychiatry*, 2000, 177:534–9.
9. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Geneva, World Health Organization, 1977.
10. *DSM-IV diagnostic and statistical manual of mental disorders*. Washington DC, American Psychiatric Association, 1994.
11. Landis JR, Kock GG. The measurement of observer agreement for categorical data. *Biometrics*, 1977, 33(1):159–74.
12. Fleitlich-Bilyk B. *The prevalence of psychiatric disorders in 7–14 year olds in the south east of Brazil* [thesis]. London, University of London, 2003.
13. Mullick MS, Goodman R. The prevalence of psychiatric disorders among 5–10 year olds in rural, urban and slum areas in Bangladesh: an exploratory study. *Social psychiatry and psychiatric epidemiology*, 2005, 40(8):663–71.
14. Thabet AA, Stretch D, Vostanis P. Child mental health problems in Arab children: application of the strengths and difficulties questionnaire. *International journal of social psychiatry*, 2000, 46(4):266–80.
15. Almaqrami MH, Shuwail AY. Validity of the self-report version of the strengths and difficulties questionnaire in Yemen. *Saudi medical journal*, 2004, 25(5):592–601.