

Validation of an Arabic multi-informant psychiatric diagnostic interview for children and adolescents: Development and Well Being Assessment-Arabic (DAWBA-Arabic)

Pia Zeinoun^{a,b}, Souha Bawab^c, Mia Atwi^a, Nayla Hariz^a, Lucy Tavitian^a, Mounir Khani^a, Ziad Nahas^a, Fadi T. Maalouf^{a,d,*}

^aDepartment of Psychiatry, American University of Beirut Medical Center (AUBMC), Beirut, Lebanon

^bSchool of Social and Behavioral Science, Tilburg University, The Netherlands

^cDepartment of Psychology, American University of Beirut (AUB), Beirut, Lebanon

^dDepartment of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Background: Countries in the Arab region lack a valid Arabic psychiatric diagnostic interview for children and adolescents. We set out to establish the diagnostic validity of the Arabic version of the Development and Well Being Assessment (DAWBA-Arabic), a multi-informant structured interview for predicting DSM-IV-TR diagnoses.

Methods: The DAWBA was translated, updated, and administered to 45 participants (child and adolescent psychiatric outpatients and their parents) as part of a clinic registry. Two clinicians, blinded to their respective diagnoses, formulated the DAWBA diagnoses. Participants also underwent a clinical evaluation by a child and adolescent psychiatrist who generated clinical diagnoses according to the DSM-IV-TR.

Results: Inter-rater reliabilities were .93, .82, and .72 for disruptive disorders, mood disorders and anxiety disorders respectively. Agreement between DAWBA and clinical diagnoses was substantial for disruptive disorders ($\kappa = .0.82$) and mood disorders ($\kappa = 0.74$), and moderate for anxiety disorders ($\kappa = 0.46$).

Conclusion: The DAWBA-Arabic could serve as a valid and reliable clinical tool for assessing psychiatric disorders among children and adolescents in the Arab region.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Arab countries lack a valid and comprehensive Arabic diagnostic instrument for assessing psychiatric disorders in children and adolescents. Existing instruments are either not adapted or not validated in Arabic yet [1,2], and those in Arabic are either not suitable for children and adolescents [3] or not sufficiently exhaustive of disorders and age-groups [4]. This gap is present despite the Arabic language being native to about 280 million speakers worldwide [5] and official in 21 countries in the Middle East and North Africa [6].

The absence of a valid Arabic instrument to assess psychiatric disorders in this age group has hindered research

on mental illness in the Arab region. First, population-based studies on the prevalence of psychiatric disorders have been absent, partially due to the lack of a valid and simultaneously cost-effective tool. In the absence of such epidemiological data, countries lack sufficient evidence to guide public policy and, mental health care planning [7], and researchers stumble to find and report reliable prevalence data about childhood and adolescent psychiatric disorders in the region. Any existing prevalence research has used widely differing methods of assessment, and shown vastly discrepant prevalence data. Second, gathering and articulating clinical data in psychiatric outpatient clinics, have also been exigent without a standardized instrument [8]. Short of a structured or semi-structured diagnostic tool, diagnoses are currently made by professionals of different levels of seniority and background, biasing the estimated prevalence of psychiatric disorders in tertiary settings. For example, a study in Oman found “hyperactivity” to be the most common presentation

* Corresponding author. Child and Adolescent Psychiatry Program, Department of Psychiatry, American University of Beirut, P.O. Box: 11-0236, Riad El Solh 11072020, Beirut-Lebanon.

E-mail address: fm38@aub.edu.lb (F.T. Maalouf).

or diagnosis in a clinical sample of children and adolescents [9], while a study in primary care settings in the United Arab Emirates found Obsessive Compulsive Disorder (OCD) as the most commonly diagnosed disorder, and Attention Deficit Hyperactivity Disorder (ADHD) as the least diagnosed [10], while mood disorders were most common in the community [11]. A standardized assessment instrument in Arabic can bring added scientific rigor to existing methods of evaluation. Such evidence-based assessment, can aid in decision-making about clinical services, especially in a time of budget constraints and limited resources.

To address this shortcoming, we aimed at expanding on an existing diagnostic instrument in Modern Standard Arabic, and establishing its diagnostic validity and interrater reliability. *The Development and Well Being Assessment* (DAWBA [12]) is a structured interview used to formulate psychiatric diagnoses of children and adolescents, ranging from age 4 to 17 years 11 months, based on the Diagnostic and Statistical Manual of Mental Disorders 4th edition, text revised (DSM-IV-TR [13]) and the International Classification of Diseases 10th edition (ICD 10 [14]). Originally developed in English, it is a multi-informant tool, consisting of a parent interview, a teacher questionnaire, and an adolescent interview for those aged 11 to 17 years 11 months. The informants respond to forced-choice questions about signs and symptoms and degree of impairment (impact), and are also allowed to freely describe the problems in open-ended questions. The information gathered is electronically processed to generate probabilities of diagnoses. Trained clinical raters subsequently review the information including symptoms, impact, open-ended comments, and computer-generated algorithm-based predictions, and either confirm computer diagnoses, reject them, or add additional diagnoses.

The DAWBA poses several advantages over other instruments. Being a structured interview that uses direct reports of respondents, it can be administered online, or by lay-interviewers with no experience in psychopathology, but sufficient training in administration. This reduces the cost of human resources in community-based and clinical studies. In addition, the DAWBA conveniently uses “skip-rules” when the child/adolescent screens negatively for specific symptoms, therefore eliminating further questioning about the disorder, and reducing administration time. Also, unique to the DAWBA is the use of open-ended questions that inquire about the main problem, its worst time, onset, duration, impact, and past treatment. These open-ended comments are paramount because they provide the clinical rater with a situational and cultural context, clarify possible discrepancies across reports, and capture symptoms that may be uncommon, but are distinct by description (e.g., trichotillomania, selective mutism). The diagnoses generated by the latest version of the DAWBA are grouped as “Emotional disorders” (subsuming all Mood Disorders and Anxiety Disorders), “Hyperactivity” (which includes all subtypes of ADHD), “Behavioral disorders” (which include Opposition-

al Defiant Disorder (ODD) and Conduct Disorder (CD), Social disorders (comprising Selective Mutism and all Attachment Disorders), and finally a category of “Others” is assigned to less common disorders such as all Pervasive Developmental Disorders (PDD), Eating Disorders, Tic Disorders, and Psychotic Disorders.

Embedded within the DAWBA, is its adjunct scale, the Strengths and Difficulties Questionnaire (SDQ [15]) — a 25-item, Likert-type, broad-band rating scale used to screen for symptoms in children aged 3 to 17 years 11 months, as reported by parents and teachers, and by adolescents aged 11 to 17 years 11 months. The SDQ includes supplemental questions on perceived impact of the problem, chronicity, distress, social impairment and burden, and inquires whether the respondent thinks that the child has a problem. The various scores yielded (Conduct Problems, Inattention–Hyperactivity, Emotional Symptoms, Peer Problems, Prosocial behavior, and Total Difficulties scores) are used in the various sections of the DAWBA interview as indicators of problem areas.

The DAWBA-English validity was initially established in a large British household study that randomly selected 471 households, and administered the DAWBA to parents, children above the age of 11, and school teachers of the adolescents [12]. Those described by parents, teachers and themselves as having a problem, were much more likely to belong to the group of participants who received a DAWBA diagnosis (38.5%, 50%, 25% respectively), than those not described as such (2.7%, 7.3%, 4.1%). The DAWBA also discriminated between British community and clinical populations. The likelihood of any DAWBA diagnosis in the clinical sample (92.3%) was much higher than that in the community sample (10.6%). Concurrent validity was supported by the finding that children and adolescents with emotional disorders, behavioral disorders or ADHD on the DAWBA had the respective SDQ scores of emotional problems, conduct problems, and hyperactivity problems, as their highest score. To investigate the extent to which DAWBA diagnoses agreed with a clinical diagnosis made blindly and independently, Goodman et al. [12] rated the DAWBA protocols of 39 clinically diagnosed outpatients from various tertiary clinical sites in Britain. Cross-tabulations revealed good agreement between the DAWBA ratings and clinical diagnoses for emotional disorders (i.e., Mood and Anxiety Disorders) and ADHD, and fair agreement for CD and ODD.

A recent study [16] compared the diagnostic agreement between the English versions of the DAWBA, Diagnostic Interview Schedule for Children (DISC [17]), and Child and Adolescent Psychiatric Assessment (CAPA [18]). Diagnostic agreement between the DAWBA and the latter two interviews was fair, with kappa ranging between 0.13 and 0.57, and mostly attributed to the methodological differences in instrument administration and rating.

The comprehensiveness yet cost-effectiveness of the DAWBA have led to its adoption in various languages,

and its usage predominantly in epidemiological studies, but also clinical settings [4,19–22]. In Bangladesh, the DAWBA-Bengali was administered to 100 children aged 5 to 17 who were referred to a psychiatric clinic [19]. The extent of agreement between the clinical diagnoses and the DAWBA diagnosis formulated by a blinded independent rater, was 0.94 for “emotional disorders”, 0.64 for “behavioral disorders”, and 0.63 for ICD-10 Hyperkinetic Disorder. It should be noted that clinical diagnosis was considered the “gold standard” since Bangladesh lacked a previously validated measure to serve as a true gold standard. Fleitlich-Bilyk and Goodman [20] established the validity of the DAWBA-Portuguese prior to its use in an epidemiological study in South East Brazil. A blinded rater generated DAWBA diagnoses for 87 clinically diagnosed patients at a mental health clinic. Results showed agreement on the primary clinical diagnosis in 78% of cases.

Evidence of the incremental usefulness of the DAWBA in clinical settings has been mixed. In Switzerland, researchers evaluated the clinical utility of the German version of the DAWBA on a sample of 270 probands and noted that the DAWBA added value to the clinical diagnostic process of internalizing disorders, but diagnostic agreement between clinician and DAWBA diagnosis was fair across diagnostic categories [22]. Another study did not find evidence that disclosure of the DAWBA diagnoses during clinical assessment helped clinicians provide more accurate diagnoses [23].

The extent of agreement between two independent blind raters of the DAWBA was established in various countries. In Brazil, inter-rater reliability on a community sample of 255 participants was excellent for “any disorder” ($\kappa = 0.93$), “internalizing” disorders ($\kappa = 0.91$), and “externalizing” disorders ($\kappa = 1.00$) [20]. The interrater agreement on 500 cases from the British Child and Adolescent Mental Health Survey of 1999, was also substantial for “any disorder” ($\kappa = 0.86$), and “externalizing disorders” ($\kappa = 0.98$), and fair for internalizing disorders ($\kappa = 0.57$) [24]. Substantial inter-rater agreement was also reported to be excellent on a sample of 60 random cases in the Swiss study. Kappa coefficients ranged from 0.79 for “Other disorder”, 0.84 for “any disorder” and “internalizing disorder”, and 0.89 for “externalizing disorder” [22].

Several sections of the DAWBA-English were adopted into Arabic in Yemen (while other sections were translated in Lebanon for this study) through translation and back-translation, preliminary piloting and subsequent changes [4]. They administered the DAWBA to 97 parents of children aged 5 to 12 across three clinical sites in Yemen. Teachers of 68 of the children also participated. The DAWBA diagnoses were generated by the primary author of the DAWBA, R. Goodman, in England, after all open-ended comments were translated from Arabic to English. The children’s clinical diagnoses were established at the clinical site by a psychiatrist, or a psychologist or a pediatrician as per DSM-IV-TR or ICD-10. For analyses,

diagnoses were collapsed into two categories of “externalizing” disorders (subsuming ADHD, ODD, CD) and “emotional disorders” which included depressive disorders, OCD, and other anxiety disorders. The DAWBA ratings and clinical diagnoses showed moderate agreement for emotional disorders ($\kappa = .634$) and externalizing disorders ($\kappa = .691$). This study provided preliminary evidence for the validity of the DAWBA on Arab children aged 5 to 12. However, there were several caveats that limited the generalizability of the Arabic DAWBA. The sample did not include children above the age of 12, and consequently did not utilize the adolescent interview of the DAWBA. Moreover, the study did not include sections on Bipolar Disorders, Substance-Related Disorders, Eating Disorders, and Psychotic Disorders.

In the aforementioned studies, comparisons of reliability and validity across societies are hindered by the lack of established equivalence across the instruments (see *discussion* for limitations when comparing instruments across cultures).

The current study aimed at expanding the DAWBA-Arabic to cover a wider breadth of disorders and age groups, refine its translation into Modern Standard Arabic (MSA) to ensure comprehensibility across various Arab countries, and establish its inter-rater reliability and diagnostic validity. We hypothesized that a) there would be a high rate of agreement on DAWBA diagnoses between two raters blinded to each other’s diagnoses and b) DAWBA-Arabic diagnoses would have a high rate of diagnostic agreement with clinical diagnoses within a psychiatric sample.

2. Methods

2.1. Adoption of the DAWBA-Arabic

The preexisting Arabic sections of the DAWBA underwent unavoidable changes in sentence structure, and format, in order to ensure linguistic comprehensibility across various Arabic dialects. For example, phrasing was simplified when possible, metaphors were avoided, and gender biased language was minimized [25]. New sections of Substance-Related Disorders, Psychotic Disorders, Eating Disorders and Bipolar Disorder, were translated from English to Modern Standard Arabic by the last author, a team of masters-level psychologists, and graduate students in psychology and public health, all fluent in both languages. MSA is the form of Arabic used in education, media broadcasts and literature [26]. Unlike spoken or colloquial Arabic which may differ considerably across countries, MSA is legible across Arab countries. The use of MSA facilitates the generalizability of the DAWBA-Arabic to other Arab-speaking countries, and avoids unfamiliar dialects and colloquialisms. Following forward translation, the sections were back-translated to English by a bilingual psychologist not familiar with the original English versions and were reviewed by the author of the original DAWBA, R. Goodman, who provided feedback on the back-translation. The team revised and finalized the Arabic sections accordingly.

2.2. Participants

A total sample of 45 children and adolescents aged between 5 and 17 years 11 months, recruited between January 2011 and February 2012 at the American University of Beirut Medical Center (AUBMC) was included in this study, out of a total of 143 referrals to the Psychiatry Department of AUBMC who participated in a clinic registry. From the total sample of 143 participants, 6 participants were excluded because the primary clinical diagnosis was suspected to be Intellectual Disability or Learning Disability and 9 participants did not return to complete the clinical evaluation. Of those remaining, 58 participants did not wish to be administered the DAWBA but completed demographic information sheet and other scales (See Fig. 1). Teacher reports were not solicited. For the purpose of this study, we only included DAWBA protocols that had all sections sufficiently complete for a diagnostic rating (a total of 45 participants of which 5 had only child version completed, 23 had only parent version completed and 17 had both parent and child versions completed).

2.3. Procedure

Informed consent was obtained from parents, and assent was obtained from children and adolescents. Participants completed the DAWBA, along with other instruments in the clinical registry on the day of their first appointment to the child psychiatry clinic, or within 2 weeks of initial appointment, before any treatment was initiated. This protocol was approved by the local Institutional Review Board (IRB). Computerized versions of the DAWBA were used. Research assistants trained in the DAWBA interviews were available to aid in the use of the computer.

2.3.1. Clinical evaluation

Clinical diagnoses for participants in the psychiatric sample were generated by a child and adolescent psychiatrist, the last author (FM), according to the DSM-IV-TR. FM was blinded to the DAWBA interview at the time the clinical diagnoses were made. Diagnoses were made following one or two visits with both parents and children seen separately

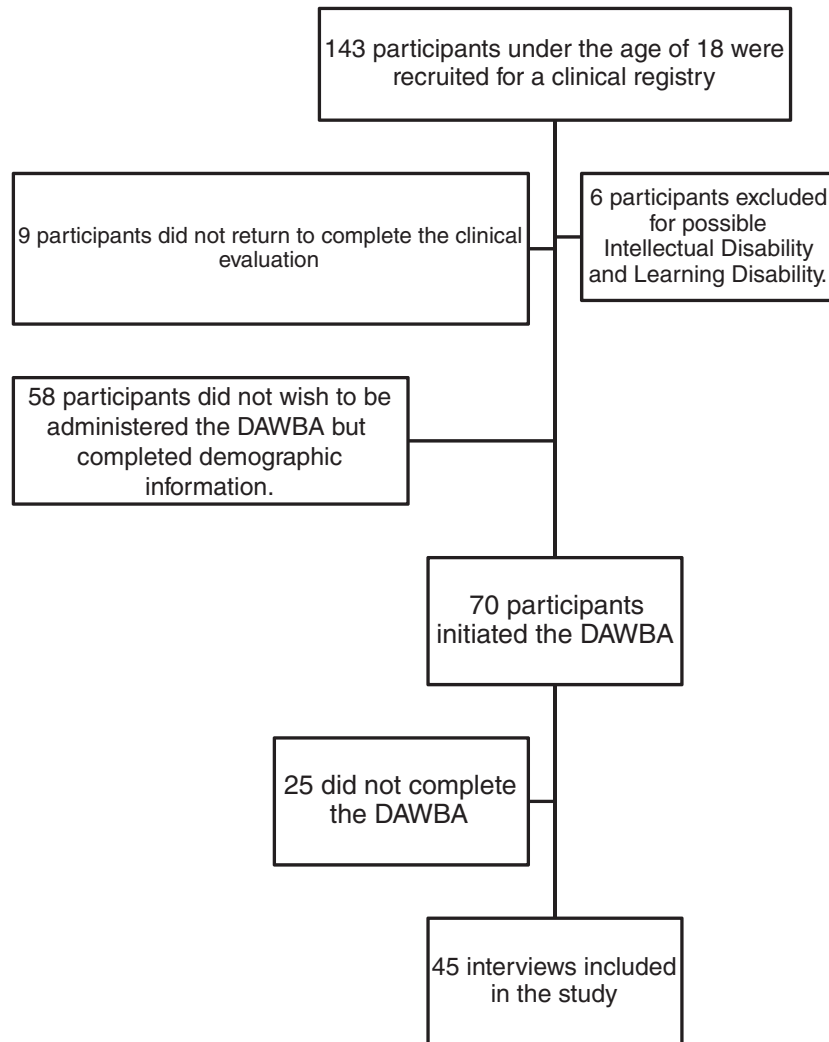


Fig. 1. Participants flow chart.

and together, for an unstructured clinical interview and observation. When relevant, the clinical evaluation utilized rating scales, review of medical and educational records, and psychological testing. Diagnoses were entered as *Mood Disorders* (subsuming Major Depressive Disorder, Bipolar Disorder, Depressive disorder NOS, Bipolar Disorder NOS), *Anxiety Disorders* (i.e. Separation Anxiety, Specific Phobia, Social Phobia, Generalized Anxiety Disorder, Obsessive Compulsive Disorder, Panic Disorder with/without Agoraphobia), *Disruptive Disorders* (Attention Deficit and Hyperactivity Disorder, Oppositional Defiant Disorder, Conduct Disorder, and Disruptive Disorder NOS), and *Other disorders* (Attachment Disorders, Pervasive Developmental Disorders, Tic Disorders, Psychotic Disorders, Eating Disorders, and Substance-Related Disorders).

2.3.2. Training of clinical raters

DAWBA-Arabic interviews were rated by two clinicians who undertook the online raters training, reviewed case studies independently, and discussed case studies to reach consensus. Some cases were also analyzed with R. Goodman, to refine the diagnostic decision-making using the DAWBA. The raters agreed a-priori to avoid the diagnosis of Not Other Specified whenever possible in order to keep interrater agreement stringent. It is arguably easier to agree that a child has some features of a disorder without specifying (i.e. NOS), than to agree on one or more specific diagnoses. Raters also agreed to keep a balance between a phenomenological approach of symptom adherence, and a culturally-sensitive approach that puts the symptoms in context. The first author (PZ), blinded to diagnoses in the psychiatric sample, and to recruitment source (psychiatric versus non-psychiatric) independently rated all DAWBA-Arabic interviews. The last author (FM) also rated a subsample of 36 anonymous DAWBA-Arabic protocols (from the first 36 participants included in this study) several months after he had seen the participants clinically, to be used only for interrater reliability.

2.4. Statistical methods

All analyses were conducted using SPSS 19. Diagnoses were entered as absent or present for each diagnostic cluster (Mood, Anxiety, Disruptive, and Other). Cohen's kappa (κ) was used to measure the degree of agreement between independent DABWA diagnoses and clinical diagnoses, as well as the level of agreement between the two DAWBA-Arabic raters, for each diagnostic category in the psychiatric sample. Kappa values greater than 0.75 represent excellent agreement beyond chance, values below 0.40 represent poor agreement beyond chance, and values between 0.40 and 0.75 represent fair to good agreement beyond chance [27,28].

3. Results

Fifty seven percent (57%) of participants ($n = 45$) were female, and 43% were male, with a mean age of 11.5 ($SD =$

4.97). There was no significant difference in age, and gender distribution between those that completed the DAWBA and those that did not ($p > .02$).

3.1. Interrater reliability

The extent to which two independent raters yield the same diagnoses from the DAWBA-Arabic interview, was evaluated in 36 cases (Table 1). The interrater reliability was excellent for disruptive disorders ($\kappa = 0.93$), and mood disorders ($\kappa = 0.82$), good for anxiety disorders ($\kappa = 0.72$), and moderate for Other Disorders ($\kappa = 0.68$).

3.2. Diagnostic validity

Distribution of diagnoses and frequencies of comorbidities according to the clinical evaluation and DAWBA-Arabic is presented in Table 2. The agreement between DAWBA and clinical diagnoses was substantial for disruptive disorders ($\kappa = 0.82$), and mood disorders ($\kappa = 0.74$). With respect to anxiety disorders the extent of agreement was fair ($\kappa = 0.46$) (Table 3).

4. Discussion

In this study we adopted the DAWBA-Arabic, by adding new sections to ensure diagnostic thoroughness, and improving translation to ensure comprehensibility across various Arabic dialects. We subsequently investigated its diagnostic reliability across raters, and its validity as a diagnostic tool in a psychiatric population. The degree of agreement between two independent, blinded DAWBA-Arabic raters was substantial for the frequently occurring Disruptive and Mood Disorders, and acceptable for Anxiety and Other Disorders. Furthermore, in a sample of children and adolescents aged 4 to 17 and 11 months, the DAWBA-Arabic ratings were substantially consistent with an independent clinical diagnosis for disruptive disorders, and mood disorders, and fairly consistent for anxiety disorders.

We found the diagnostic agreement on Disruptive Disorders ($\kappa = .82$) to be higher than that reported in previous studies including the one that used the older Arabic version of DAWBA [4,19]. We cannot compare the diagnostic agreement for Mood Disorders or Anxiety Disorders with other studies, since this is the only study that used narrow and rigorous diagnostic categories, as opposed to previous studies that often collapsed Anxiety and Mood disorders into "Emotional Disorders".

Table 1
Inter-rater reliability of DAWBA-Arabic diagnosis ($n = 36$).

Diagnosis	Kappa (Standard Error)
Disruptive Disorders	0.93 (0.06)
Mood Disorders	0.82 (0.10)
Anxiety Disorders	0.72 (0.09)
Other Disorders	0.68 (0.16)

Table 2
Frequencies of Diagnoses and Comorbidities in the Sample (n = 45).

DSM-IV TR Diagnosis	DAWBA-Arabic ^a	Clinical ^b
Disruptive Disorder	18/45 (40%)	20/45 (44.4%)
Mood Disorder	14/45 (31.1%)	13/45 (28.9%)
Anxiety Disorder	19/45 (42%)	14/45 (31.1%)
Other	6/45 (13.3)	13/45 (28.9%)
Any Diagnosis	42/45 (93%)	45/45 (100%)
No Diagnosis	3/45 (6%)	0/45 (0.0%)
One Diagnosis	27/45 (60%)	31/45 (68.8%)
Two Diagnoses	15/45 (33%)	13/45 (28.8%)
Three Diagnoses	0/45 (0%)	1/45 (2.2%)

^a Total number of diagnoses based on the DAWBA-Arabic = 57.

^b Total number of diagnoses based on clinical evaluation = 60.

The DAWBA-Arabic tended to yield more diagnoses of anxiety disorders than the clinician, resulting in fair diagnostic agreement for this group of disorders. Similar findings of fair agreement between DAWBA and clinical diagnoses have been recently reported in the German version of the DAWBA [22]. Anxiety was also the least agreed-upon diagnosis across the DAWBA the DISC and the CAPA interviews, although the DAWBA did not yield less diagnoses than the other interviews [16]. Furthermore, low to moderate agreement for narrow (as opposed to broad) diagnostic categories have been reported by other studies comparing agreement between structured interviews and clinical evaluations [29,30]. There are several possible explanations for the fair diagnostic agreement on anxiety disorders, as compared to the substantial agreement for disruptive disorders and mood disorders in this study. First, most anxiety diagnoses by the DAWBA-Arabic were comorbidities, not primary diagnoses. This may imply that during the psychiatric evaluation, comorbid anxiety disorders were under-diagnosed because the clinician and/or the patient/parent focused on the primary diagnosis (e.g. major depressive disorder). Second, patients may have been more likely to report anxiety

symptoms in a structured interview rather than in a clinical evaluation especially that the DAWBA asks in extensive details about anxiety symptoms in 7 different sections. Lastly, the difference between the number of informants used for the clinical interview and DAWBA may partially explain this discrepancy.

The degree of diagnostic agreement between a psychiatric evaluation and a DAWBA-Arabic diagnosis was more elevated in this study than in the initial validation study of the DAWBA in Yemen [4]. This may be due to the fact that, DAWBA-Arabic diagnoses were formulated by Arab psychologists and psychiatrists who are likely to be more familiar with the Arab and Lebanese culture, and hence appreciate the cultural underpinning of the open-ended comments in their native language. Studies that utilize translators or raters, who are very experienced in the DAWBA but not entrenched with the local culture, may yield less agreement with a clinical evaluation. The improved diagnostic agreement may also be due to consistency in clinical evaluations, since all diagnoses were conducted by one child and adolescent psychiatrist, as opposed to various professionals of diverse levels of seniority and expertise.

Possible limitations of this study are the relatively small sample size, and possibility of a sample bias. A large percentage of participants either refused to fill out the DAWBA or did not complete it. No difference between these 2 groups was observed however on main demographic variables.

In addition, importing instruments developed in Western country, to a linguistically and culturally differing target population, involves several limitations or biases on the level of constructs, method and items [31]. Although the DAWBA-Arabic underwent meticulous revisions by a team of bilingual mental health professionals knowledgeable with the local culture, language, and instrument, there are several possible biases inherent in the DAWBA-Arabic.

On the level of construct, the DAWBA sections and items were not adapted to cover indigenous conceptions of mental illness, or additional emic constructs. Therefore the DAWBA-Arabic, like most adopted etic instruments, has the disadvantage of insufficiently covering indigenous constructs. However, imported measures can be a meaningful source of data [32], and the DAWBA-Arabic in particular may be more culturally-sensitive due to its usage of open-ended questions. These verbatim responses allow the rater to “listen” to the participant, and hence obtain rich information about indigenous expressions of disorders. Raters, who are experienced with the local culture, can appreciate these culture-specific presentations when rejecting or accepting a diagnosis. For instance, adolescents and their parents in this study tended to endorse symptoms of separation anxiety more so than what is reported in the literature or diagnosed clinically. Raters remained aware of the cultural influences that impacted these reports — families in Arab countries are a central unit to the adolescent’s life, parents and adolescents maintain close physical proximity, sleepovers are not

Table 3
Agreement of clinical diagnosis and DAWBA-Arabic diagnosis.

Clinical DSM-IV-TR Diagnosis	DAWBA-Arabic DSM-IV TR Diagnosis		Total	Kappa (SE)
	No	Yes		
Disruptive Disorders				
No	24	1	25	
Yes	3	17	20	
Total	27	18	45	0.82 (0.08)
Mood Disorders				
No	29	3	32	
Yes	2	11	13	
Total	31	14	45	0.74 (0.11)
Anxiety Disorders				
No	24	7	31	
Yes	4	10	14	
Total	28	17	45	0.46 (0.13)

common, and children are expected to depend on their parents until adulthood. Therefore, reported symptoms of sleeping near parents, being anxious about separating and not wanting to sleep outside the house, had to be understood in this context. The open-ended question aided the raters in understanding whether the behaviors reported were pathological or culturally-sanctioned. Future studies can use combined emic and etic approaches to develop more culturally-sensitive instruments.

Method bias is a source of bias attributed to the administration procedure of the measurement [31]. Possible biases in the administration of the DAWBA-Arabic may have been introduced from the use of a computer and the administration of the DAWBA after the clinical interview with the psychiatrist (priming effect). To minimize this bias, research assistants were available to help with the computer use, and explain items if needed. Priming effects were reduced by administering a number of DAWBA interviews prior to any clinical evaluation.

Item bias can result from factors such as poor wording, inaccurate translations, and inappropriateness of item content in a cultural group [33]. Item bias was qualitatively addressed and minimized through forward translation by bilinguals, team meetings to reach consensus, and the use of culturally-neutral examples [25]. Although further reduction of item bias may have been accomplished by an independent back translation of *all* the sections, by someone not familiar with the instrument [34] the size of the DAWBA-Arabic interview made this logistically not viable.

5. Conclusion

As the disciplines of psychiatry and psychology continue to be penetrated by evidence-based practice, the need for standardized diagnostic assessments, clinically and in research, is on the rise. Mental health professionals today are required to justify diagnoses, treatments, and admissions, as the budget constraints become more palpable. DAWBA-Arabic provides an opportunity for the Arab world to introduce standardized assessments of children and adolescents into clinical practice, and consequently inform treatment, clinical policy and public policy. In addition, the current research gap in community-based studies on mental health in Arab-speaking children and adolescents can begin to mend, as the DAWBA-Arabic can be utilized in large epidemiological studies.

Acknowledgment

The authors would like to thank Dr. Robert Goodman for his valuable input during the implementation of this study and for being available to consult on rating cases when needed.

References

- [1] Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 1997;36:980-8.
- [2] Reich W. Diagnostic Interview for Children and Adolescents (DICA). *J Am Acad Child Adolesc Psychiatry* 2000;39:59-66.
- [3] Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The composite international diagnostic interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1988;45:1069-77.
- [4] Alyahri A, Goodman R. Validation of the Arabic strengths and difficulties questionnaire and the development and well-being assessment. *East Mediterr Health J* 2006;12:138-46.
- [5] Procházka S. Arabic. In: Brown K, editor. *Encyclopedia of language and linguistics*. Amsterdam, the Netherlands: Elsevier Ltd; 2006. p. 423-31.
- [6] Maamouri M. Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region. *The Mediterranean Development Forum*. The World Bank: Washington, DC; 1998.
- [7] Al-Krenawi A. Mental health practice in Arab countries. *Curr Opin Psychiatry* 2005;18:560-4.
- [8] Dreyer NA, Garner S. Registries for robust evidence. *JAMA* 2009;302:790-1.
- [9] Al-Sharbaty MM, Al-Hussaini AA, Antony SX. Profile of child and adolescent psychiatry in Oman. *Saudi Med J* 2003;24:391-5.
- [10] Eapen V, Al-Sabosy M, Saeed M, Sabri S. Child psychiatric disorders in a primary care Arab population. *Int J Psychiatry Med* 2004;34:51-60.
- [11] Eapen V, Jakka ME, Abou-Saleh MT. Children with psychiatric disorders: the Al Ain Community Psychiatric Survey. *Can J Psychiatry* 2003;48:402-7.
- [12] Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000;41:645-55.
- [13] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed., text revision. Washington, DC: American Psychiatric Publishing; 2000.
- [14] World Health Organization. *ICD-10: international statistical classification of diseases and related health problems*. 10th rev ed. New York, NY: World Health Organization; 2008.
- [15] Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997;38:581-6.
- [16] Angold A, Erkanli A, Copeland W, Goodman R, Fisher PW, Costello EJ. Psychiatric diagnostic interviews for children and adolescents: a comparative study. *J Am Acad Child Adolesc Psychiatry* 2012;51:506-17.
- [17] Costello EJ, Edelbrock CS, Duncan MK, Kalas R. Testing of the NIMH Diagnostic Interview Schedule for Children (DISC) in a clinical population: final report to the Center for Epidemiological Studies. NIMH: Pittsburgh, University of Pittsburgh; 1984.
- [18] Angold A, Prendergast M, Cox A, Harrington R, Simonoff E, Rutter M. The child and adolescent psychiatric assessment (CAPA). *Psychol Med* 1995;25:739-54.
- [19] 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. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:663-71.
- [20] Fleitlich-Bilyk B, Goodman R. Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *J Am Acad Child Adolesc Psychiatry* 2004;43:727-34.
- [21] Goodman R, Slobodskaya H, Knyazev G. Russian child mental health — a cross-sectional study of prevalence and risk factors. *Eur Child Adolesc Psychiatry* 2004;14:28-33.

- [22] Aebi M, Kuhn C, Winkler Metzke C, Stringaris A, Goodman R, Steinhausen HC. The use of the development and well-being assessment (DAWBA) in clinical practice: a randomized trial. *Eur Child Adolesc Psychiatry* 2012;21:559-67.
- [23] Ford T, Last A, Henley W, Norman S, Guglani S, Kelesidi K, et al. Can standardized diagnostic assessment be a useful adjunct to clinical assessment in child mental health services? A randomized controlled trial of disclosure of the Development and Well-Being Assessment to practitioners. *Soc Psychiatry Psychiatr Epidemiol* 2013; 48:583-93.
- [24] Ford T, Goodman R, Meltzer H. The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry* 2003;42:1203-11.
- [25] Brislin RW. Translation and content analysis of oral and written materials. In: Triandis HC, Berry JW, editors. *Handbook of cross-cultural psychology*. Boston, MA: Allyn & Bacon; 2003. p. 389-444.
- [26] Al-Tamimi FY. Assessment of language development in Arabic. *Encyclopedia of Language and Literacy Development*. Canadian Language and Literacy Research Network. London, ON. 2011.
- [27] Fleiss JL, Levin B, Paik MC. The measurement of interrater agreement. *Statistical methods for rates and proportions*. Hoboken, NJ: Wiley; 2004. p. 598-626.
- [28] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- [29] Ezpeleta L, de la Osa N, Doménech JM, Navarro JB, Losilla JM, Júdez J. Diagnostic agreement between clinicians and the Diagnostic Interview for Children and Adolescents – DICA-R – in an outpatient sample. *J Child Psychol Psychiatry* 1997;38:431-40.
- [30] Carlson GA, Kashani JH, de Fatima Thomas M, Vaidya A, Daniel AE. Comparison of two structured interviews on a psychiatrically hospitalized population of children. *J Am Acad Child Adolesc Psychiatry* 1987;26:645-8.
- [31] van de Vijver FJR, Leung K. *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage; 1997.
- [32] Church T. Prospects for an integrated trait and cultural psychology. *Eur J Pers* 2009;23:153-82.
- [33] van de Vijver FJR, Poortinga Y. Conceptual and methodological issues in adapting tests. In: Hambleton RK, Merenda PF, Spielberger CD, editors. *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005. p. 39-63.
- [34] Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess* 1994;6:304-12.