The acceptance of the K-SADS-PL – potential predictors for the overall satisfaction of parents and interviewers

TINA MATUSCHEK,^{1,2} SONIA JAEGER,^{1,2} STEPHANIE STADELMANN,^{1,2} KATRIN DÖLLING,¹
STEFFI WEIS,^{1,2} KAI VON KLITZING,^{1,2} MADLEN GRUNEWALD,^{1,2} ANDREAS HIEMISCH¹ &
MIRKO DÖHNERT^{1,2}

1 LIFE – Leipzig Research Centre for Civilization Diseases, Universität Leipzig, Leipzig, Germany

2 Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, Universität Leipzig, Leipzig, Germany

Key words

K-SADS-PL, acceptance, child psychiatry, research setting, multilevel modelling

Correspondence

Tina Matuschek, LIFE – Leipzig Research Centre for Civilization Diseases, Universität Leipzig, Philipp-Rosenthal-Straße 27, 04103 Leipzig, Germany.
Email: tina.matuschek@life.uni-leipzig.de

Received 1 July 2014;
revised 28 October 2014;
accepted 14 November 2014

Abstract

The presented study investigated the interviewee (parents) and interviewer acceptance of the semi-structured diagnostic interview Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children Present Lifetime version (KSADS-PL; German version). Seventeen certified interviewers conducted 231 interviews (interviewers conducted several interviews; interviewees were only questioned once). Interviewees and interviewers anonymously rated their acceptance right after the interview was finished. The nested data structure was analysed regarding an individual interviewer bias and potential predictors of overall satisfaction. Therefore, factors improvable by interviewer training were included, as well as fixed factors which cannot be improved by professional training. The overall satisfaction was evaluated as highly positive with significant higher interviewee and interviewer ratings in the research as compared to the clinical recruitment setting. An individual bias of the interviewer on his or her own acceptance over time, but not on the evaluation of the corresponding interviewee was found. Neither the professional background nor the gender of the interviewer had a significant contribution in predicting these differences. The interviewer model showed no significant change over time and only the interview duration and the interviewee acceptance were significant predictors for interviewer overall satisfaction. Regarding the interviewee model, just the interviewer acceptance was a significant predictor. Copyright © 2015 John Wiley & Sons, Ltd.

Introduction

Semi-structured or structured diagnostic interviews are the gold standard in diagnosing psychiatric disorders. Besides the Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children (6–18 years) Present-Lifetime version (K-SADS-PL; Kaufman *et al.*, 1996) – which is of central interest in this study – a wide number of different diagnostic interviews exists [e.g. the Preschool Age Psychiatric Assessment (PAPA; Egger *et al.*, 1999) and the National Institute of Mental Health Diagnostic Interview Schedule for Children (NIMH DISC-IV; Shaffer *et al.*, 2000)]. Even though their usefulness is widely known, diagnostic interviews are still insufficiently used in clinical settings because of concerns about their acceptance and feasibility (Ehlert, 2007; Joiner *et al.*, 2005; Schneider and Döpfner, 2004; Silverman and Ollendick, 2005). Accordingly, clinicians often rely more on their clinical judgment and their experience, though this may lead to unreliable diagnoses (Beck *et al.*, 1962; Jensen and Weisz, 2002; Pinninti *et al.*, 2003; Shear *et al.*, 2000). Studies evaluating interview acceptance showed, on the one hand, that most interviewers value that diagnostic interviews are more detailed and serve to get a complete picture of the symptomatology. On the other hand, interviewers think that diagnostic interviews are more time consuming than unstructured exploration. The interviewees evaluated diagnostic interviews as mostly positive according to a number of different studies (Bruchmüller *et al.*, 2011; Helzer, 1981; Marshall *et al.*, 2001; Pinninti *et al.*, 2003; Saile *et al.*, 2000; Sørensen *et al.*, 2007; Suppiger *et al.*, 2009; Zahner, 1991). Due to a very heterogeneous pool of instruments for evaluating the interview acceptance, it is difficult to compare results between studies and different interviews. Most of the former mentioned studies used questions without differentiated and consistent response categories. An exception is the study by Suppiger *et al.* (2009). They used structured parallel versions of interviewee and interviewer questionnaires with Likert-type scales to analyse the acceptance of the Diagnostisches Interview für psychische Störungen (DIPS; Margraf *et al.*, 1994).

As far as we know, only one pilot study evaluating the acceptance of the K-SADS-PL exists (Danish interview version; Sørensen *et al.*, 2007). In this study, a parent and a child interview version was used and both, parents and children, had to evaluate their interview satisfaction. All interviews were conducted by only one interviewer in a clinical setting. The psychiatric staff was asked about the usefulness of the interview. In general, the results showed a high parental satisfaction with the K-SADS-PL.

Parents had some concerns in using the interview for young children with conduct disorders, because some of the children were more sad, hyperactive or difficult after the interview. The staff described the K-SADS-PL as a predominantly useful diagnostic tool.

Our aim was to evaluate the interviewee and interviewer acceptance of the K-SADS-PL (German version) in a clinical and a research setting (Delmo *et al.*, 2001). Our major goals were first, to replicate the results on parental acceptance found in the pilot study by Sørensen *et al.* (2007) using more differentiated questionnaires and a larger sample size. For a better comparability between results of different studies on acceptance, we went for adapted versions of the acceptance questionnaires by Suppiger *et al.* (2009). In line with the aforementioned pilot study, we hypothesized a high overall satisfaction of the interviewees. Second, we included different interviewers in our analyses. Here, we expected a high overall satisfaction, too. Third, we hypothesized higher overall satisfaction ratings of interviewees and interviewers in the research recruitment setting, as shown by Suppiger *et al.* (2009) for the DIPS.

Furthermore, as interviewers conducted several interviews combined with acceptance ratings we expected to find an individual interviewer bias. This is based upon the consideration that inherent individual features (e.g. professional background, personality) of the interviewer may influence the personal perception and evaluation of both interview partners in the interview situation. It may also lead to other dependencies in the data, e.g. all parents interviewed by the same experienced interviewer may have similar interview durations. Therefore, we analysed the data with regard to their nested structure. Additionally, we were interested in an in depth analysis of potential influencing factors on the acceptance of both interview partners. We considered two kinds of factors: factors that can be improved by professional interviewer training (interview duration, the experience in conducting the K-SADS-PL, acceptance of the interview partner) and fixed factors which cannot be improved by professional training, but which might have had an effect on the acceptance rating, too (level of current psychopathological severity and impairment of the child, educational level and recruitment setting of the interviewee, professional background and gender of the interviewer).

Method

Sample

The data reported here were collected from May 2011 until February 2013 and resulted in a sample of 231

interviews for which at least one acceptance questionnaire was completed by either the interviewee or the interviewer (for $n = 218$ both acceptance questionnaires exist).

Interviewees

Interviewees were parents participating in the research project LIFE Child Depression with one of their children. The families were recruited through the Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, Universität Leipzig (clinical setting) as well as through a cooperating research project called LIFE Child Health [for a detailed description of the study see Quante *et al.* (2012)], and through additional recruitment procedures (research setting). Families with a child between age eight years zero months and 14 years 11 months could take part in the study. The parent version of the K-SADS-PL was used. The majority of interviews were conducted with the biological or social mother (84.8% of the interviews). Other interviewees were, for example, the biological or social father (4.8%) and others (8.6%; e.g. both parents). Of these 51.9% of the children were diagnosed with at least one present diagnose. Specifically, 21.7% of the children ($n = 50$) were diagnosed with a conduct disorder, 20.4% ($n = 47$) with at least one anxiety disorder, 20.4% ($n = 47$) with attention deficit hyperactivity disorder (ADHD), 19.5% ($n = 45$) with at least one other diagnose (e.g. tic disorder) and 8.7% ($n = 20$) with an affective disorder.

The educational level of the interviewees was summarized into three categories (Gall *et al.*, 2010): low, including all schooling up to the General Certificate of Secondary Education; intermediate, including the diploma from German secondary school qualifying for university admission or matriculation (Abitur), trades and apprenticeships; and high, including bachelor degrees and higher. The results are shown in Table 1. A significant contingency was found between the recruitment setting and the educational level of the interviewees, $\chi^2(2) = 22.46$, $p < 0.001$.

Interviewers

Because of the implementation of the K-SADS-PL in the clinical setting and in the research project, a high number of interviewers was involved ($n = 17$; three male; 14 female). All interviewers had a professional background, being psychologists ($n = 6$), master-students in psychology ($n = 6$), or junior doctors in training for child and adolescent psychiatry ($n = 5$). Seven interviewers had primarily a clinical background and nine interviewers were part of the research staff. All interviewers had to do an intensive standardized training before being allowed to conduct

Table 1. Educational level of the interviewees (in per cent) comparing the two recruitment settings

Educational level	Clinical setting ($n = 115$)	Research setting ($n = 116$)
Low ¹	9.6	2.6
Intermediate ²	73.0	57.8
High ³	11.3	37.1
No information	6.1	2.6

¹All schooling up to the General Certificate of Secondary Education.

²Diploma from German secondary school qualifying for university admission or matriculation (Abitur), trades and apprenticeships.

³Bachelor degrees and higher.

interviews on their own. The training comprised of private studying of the K-SADS-PL manual, a theoretical instruction to the interview, several group exercises, and assistance to several interviews of experienced interviewers. Finally, all interviewers had to do several interviews on their own until they met required quality standards. The whole training was guided by an experienced and well-trained interviewer.

Additionally, for all interviewers a permanent consulting service was obligatory. Furthermore, all interviews were checked for correctness to assure quality in the conclusion and documentation of diagnoses. The inter-rater reliability [$n = 23$ interviews (10%) were double rated] was on a substantial level (Kappa = 0.71) for conducted diagnoses and on a high level [intra-class correlation coefficient (ICC) = 0.93] for the Global Assessment Scale of Functioning [GAF; American Psychiatric Association (APA), 2000; Saß *et al.*, 2003].

Measures

K-SADS-PL

The K-SADS-PL is an internationally well-known and widely used semi-structured diagnostic interview to assess present and lifetime psychopathology in children and adolescents diagnosing psychiatric disorders and subsyndromal symptomatology, according to DSM-IV-TR (APA, 2000; Saß *et al.*, 2003) and ICD-10 (World Health Organization, 1994). It consists of screening parts with key symptoms for each disorder and supplemental diagnostic assessments (Kaufman *et al.*, 1996). Thirty psychiatric disorders can be assessed. Additionally, the GAF can be evaluated after conducting the interview. GAF is a scale to rate the current

general functioning (psychological, social and school-related) of the child [modified version of the Global Assessment Scale (GAS; Endicott *et al.*, 1976) which is a revised version of a scale by Luborsky (1962)]. Some changes were made to the German version of the K-SADS-PL (Delmo *et al.*, 2001) mostly according to the K-SADS-PL 2009 working draft (Axelson *et al.*, 2009). Furthermore, we paraphrased all questions into a parent version.

Acceptance questionnaires

We used two questionnaires by Suppiger *et al.* (2009) which we adapted for our purposes. For example, all items had to be paraphrased into a parent version. Both questionnaires started with a question about the overall satisfaction with the interview, which was rated on a scale from “not at all satisfied” (=0) to “totally satisfied” (=100). This was followed by 10 items representing the mental performance (e.g. interviewer version: “I conducted the interview as well as I could”) and emotional reaction (e.g. interviewee version: “The relationship to the interviewer was positive”) of the interviewee/interviewer during the interview. Items were rated on a four-point scale (0 = disagree, 1 = slightly agree, 2 = almost completely agree, 3 = completely agree).

Procedure

For families recruited through the psychiatric clinic, the K-SADS-PL was part of the standard diagnostic process and took place in a clinical setting. If they gave written consent to take part in the research project after detailed information about all parts of the study, the K-SADS-PL data were used for research purposes. Families recruited through other ways had a combined appointment in the LIFE study centre where they also received detailed information. After giving written consent, they were interviewed. In most cases, the interview could be finished during one appointment in both recruitment settings. Right after it was finished, the interviewee and the interviewer anonymously answered their acceptance questionnaire, put it into a coded envelope and sealed it to secure anonymity. Finally, the interviewer coded DSM-IV-TR disorders, subsyndromal symptoms and the GAF based on the conducted interview. All interviews were audiotaped in order to facilitate the analyses of diagnoses and to assess inter-rater reliability.

Statistical analysis

First, descriptive analyses, bivariate analyses [correlation coefficients according to Spearman's ρ (r)], and

inferential statistics (Mann–Whitney U test) comparing the two recruitment settings were calculated for the interviewee and the interviewer acceptance questionnaire, respectively. Second, descriptive statistics on the number of conducted interviews, the interview duration, the GAF, and the number of diagnoses were calculated regarding the entire sample (all settings) and comparing the two different recruitment settings. These statistical analyses were conducted with IBM SPSS Statistics 20.

Finally, two multilevel models were tested to analyse the nested data structure and to look for potential predictors of interviewer and interviewee overall satisfaction (Hierarchical Linear and Nonlinear Modeling, HLM 7; Raudenbush *et al.*, 2011). In each of the models we hypothesized two levels, taking into account that interviewees were nested within interviewers. Level 1 thus constitutes ratings concerning a particular interview, whereas on Level 2 differences between the different interviewers were taken into account. Regarding the model on overall satisfaction of the interviewee, Level 1 units of analysis were the ratings of the interviewees. Regarding the model on overall satisfaction of the interviewer, Level 1 units of analysis were the interviewer ratings of a particular interview, thus constituting repeated measures on interviewer acceptance (within-person equations). The predictors included into both multilevel analyses on Level 1 were: the interview duration (in minutes), the educational level and the recruitment setting (clinical versus research setting) of the interviewee, the GAF, the overall satisfaction rating of the interview partner and the experience in conducting the K-SADS-PL. In the interviewee model, the experience in conducting the K-SADS-PL was operationalized through the number of interviews an interviewer had conducted until this particular interview (i.e. zero if it was the first interview, one if it was the second, ...). In the interviewer model, this predictor was transformed into the basic time parameter, with the first interview establishing the reference point. Additionally on Level 2, the professional background of the interviewer (predominant clinical versus scientific work experience) and the gender of the interviewer were added as possible predictors of Level 1 effects. The restricted maximum likelihood approach was used. The intercept-coefficient was entered as “random”, to test for hypothesized intercept variation between interviewers and interviewee groups, respectively. The slope-coefficients of all predictors were entered as “fixed” in both models, assuming to be the same for all groups.

Results

Descriptive results on predictors and number of diagnoses

Table 2 gives an overview of means (*M* values) and standard deviations (SDs) of the numbers of conducted interviews per interviewer, the interview duration, the GAF, and the number of diagnoses (present and lifetime) regarding the entire sample (all settings) and comparing the two recruitment settings (clinical and research setting).

Overall, a weak negative correlation was found between the numbers of conducted interviews and the interview duration ($r = -0.21$; $p = 0.002$), and a moderate negative correlation between the GAF and the interview duration ($r = -0.49$; $p < 0.001$). The overall satisfaction with the interview was rated highly positive by interviewers and interviewees in all settings with a moderate correlation between the overall satisfaction ratings of interviewers and interviewees ($r = 0.31$; $p < 0.001$). On that high level, interviewers rated their overall satisfaction more conservative than did the interviewees.

Interviewee acceptance

In total, 227 interviewee acceptance questionnaires were completed. The mean overall satisfaction of the interviewees was $M = 90.69$ (SD = 12.44) with a range from

30 ($n = 1$) to 100 ($n = 100$) (see Supplement 1 in the Supplementary Material for means and standard deviations of all questionnaire items). Comparing the overall satisfaction of the interviewees in the clinical setting (median = 90.00) and the research setting (median = 100.00) a significant difference was found, $U = 4100.50$, $z = -2.57$, $p = 0.01$, $r = -0.18$. Interviewees of the research setting had a significantly higher overall satisfaction than interviewees of the clinical setting. Individual correlations between the questionnaire items, the overall satisfaction and the predictors are presented in Supplement 2.

Interviewer acceptance

Analysing the 222 completed interviewer questionnaires, a mean overall satisfaction of $M = 81.76$ (SD = 13.98) with a range from 20 ($n = 1$) to 100 ($n = 34$) was found (see Supplement 3 for means and standard deviations of all questionnaire items). The overall satisfaction of the interviewers in the research setting (median = 90.00) was significantly higher than the overall satisfaction of the interviewers in the clinical setting (median = 80.00), $U = 4637.00$, $z = -3.28$, $p < 0.001$, $r = -0.22$. Furthermore, there were significant correlations between the overall satisfaction and the numbers of interviews conducted ($r = 0.30$; $p < 0.001$), the interview duration ($r = -0.34$; $p < 0.001$) and the GAF ($r = 0.26$; $p < 0.001$) (see Supplement 4 for individual correlations between the questionnaire items, the overall satisfaction and the predictors).

Multilevel models

Interviewee

We started with an intercept only model to test for significant mean variation in the overall satisfaction of the interviewees between interviewer groups, as well as residual variance indicating unexplained variation between interviewees (χ^2 -test). The deviance of this model with two parameters ($\text{no}_{\text{parameters}} = 2$) was 1587.27.

As the ICC (ICC = 0.004; $\chi^2(16) = 13.44$, $p > 0.5$) and the reliability estimate = 0.045 indicated, there was no significant overall mean variation between the groups, but sufficient residual variance within them. Therefore, the intercept-coefficient was entered as “fixed” in further analyses. Next, we added all hypothesized predictors to explain the variance on Level 1. As Table 3 shows, only the overall satisfaction of the interview partner (interviewer) accounted for a significant proportion of variance in the model and was a significant predictor for the overall satisfaction of the interviewee.

Table 2. Mean (*M*) and standard deviation (SD) of number of conducted interviews, interview duration (in minutes), GAF scale, and number of diagnoses for the entire sample (all settings) and different settings

	All settings (<i>N</i> = 231)	Clinical setting (<i>n</i> = 115)	Research setting (<i>n</i> = 116)
	<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)
Number of interviews	10.92 (8.47)	9.14 (7.7)	12.69 (8.86)
Interview duration	99.87 (52.66)	116.79 (59.45)	84.59 (40.18)
GAF ¹	67.40 (17.36)	55.35 (12.51)	79.03 (12.88)
Number of diagnoses (present)	1.04 (1.30)	1.77 (1.36)	0.31 (0.68)
Number of diagnoses (lifetime)	1.48 (1.72)	2.39 (1.83)	0.59 (0.98)

¹GAF rated on a scale of 0 to 100.

Table 3. Beta values (β), standard errors of beta values ($SE\beta$), and significance values for the multiple regression analysis with overall satisfaction of the interviewee as dependent variable

	β	$SE\beta$
<i>Intercept</i>	84.21***	3.87
<i>Level 1-predictors</i>		
Number of interviews	−0.01	0.11
Interview duration	0.01	0.02
GAF	−0.01	0.08
Overall satisfaction		
interview partner (interviewer)	0.37***	0.08
Recruitment setting	1.34	2.68
Educational level interviewee		
Intermediate	5.46	3.93
High	5.85	4.36

* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$.

Interviewer

Regarding the overall satisfaction of the interviewer, we first calculated an intercept only model to test for an overall mean variation between interviewers, as well as residual variance indicating unexplained within-person variation over time (χ^2 -test). The deviance of this model ($no_{parameters} = 2$) was 1778.83. As the ICC ($ICC = 0.33$; $\chi^2(16) = 62.26$, $p < 0.001$) and the reliability estimate = 0.68 showed, there was a significant overall mean variation between interviewers and a sufficient residual variance over time.

Next, we added a time variable, allowing for linear change in the overall satisfaction of the interviewer over time. Furthermore, we added all hypothesized time-variant predictors on Level 1 and two time-invariant variables (professional background and gender of interviewer) on Level 2. As Table 4 shows, no significant linear change over time was found. The interview duration and the overall satisfaction of the interview partner (interviewee) were significant time-variant predictors for the overall satisfaction of the interviewer. Professional background and gender of the interviewer were no significant predictors of the identified intercept-variation in the overall satisfaction. Adding all described predictors improved the model fit, $\chi^2(14) = 75.83$, $p < 0.001$. The deviance of this model ($no_{parameters} = 2$) was 1265.28.

Table 4. Beta values (β), standard errors of beta values ($SE\beta$), and significance values for the multiple regression analysis with overall satisfaction of the interviewer as dependent variable

	β	$SE\beta$
<i>Intercept</i>	76.92***	6.31
Professional background	1.62	4.98
Sex	−0.66	5.97
<i>Time</i>	0.005	0.11
<i>Level 1-predictors</i>		
Interview duration	−0.07***	0.02
GAF	−0.05	0.07
Overall satisfaction interview	0.26***	0.06
partner (interviewee)		
Recruitment setting	0.72	2.49
Educational level interviewee		
Intermediate	3.77	3.41
High	6.6	3.76

* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$.

Discussion

According to the presented results and in line with the findings of Sørensen *et al.* (2007), the K-SADS-PL can be regarded as a highly accepted diagnostic interview in different settings. Furthermore, it seems to be feasible for a wide range of psychopathology, from no clinically relevant symptoms (about half of the children in our study had no present diagnose) to multiple comorbidities of psychiatric disorders.

As far as we know, our study is the first examining the acceptance of the K-SADS-PL from two important parallel viewpoints: interviewees and interviewers. Moreover, we included male and female interviewers with different professional background and work experience in our study. Additionally, potential predictors of the overall satisfaction were examined, taking the nested data structure into account. Here, we distinguished between variables that can be improved by professional training of the interviewers (e.g. experience in conducting the K-SADS-PL) and variables that ought to be unswayable by that (e.g. GAF).

As hypothesized, the overall satisfaction with the interview was evaluated highly positive by interviewers and interviewees in both settings. In line with results by Suppiger *et al.* (2009), we found significant higher overall satisfaction of both interview partners in the research recruitment

setting as compared to the clinical setting. Nevertheless, the overall satisfaction in the clinical setting was also high, the differences between the two settings were small, and the recruitment setting was no significant predictor in both multilevel models.

Regarding the multilevel model on interviewee overall satisfaction, no nested data structure with inter-individual differences between the interviewers was found. Accordingly, it did not matter which interviewer conducted the interview. Furthermore, the interviewee satisfaction seemed to be quite robust against the influence of setting variables such as the interview duration or the experience of the interviewer. Only the acceptance of the interview partner (interviewer) was a significant predictor of the interviewee satisfaction.

Regarding the multilevel model on interviewer overall satisfaction, a nested data structure was found. The inter-individual differences between the interviewers indicate that interviewers have an individual bias (level) in rating their satisfaction over time. Neither the professional background nor the gender of the interviewer had a significant contribution in predicting these differences. There may be other important predictors that explain this effect, such as personality traits of the interviewer. Furthermore, the interviewer model showed no significant change over time. Thus, the experience in conducting the K-SADS-PL had no direct influence on the overall satisfaction. Only the interview duration and the acceptance of the interview partner (interviewee) were significant predictors.

There are some limitations to our study. First, we exclusively used the parent interview version. Thus, conclusions about the interview acceptance are limited to the parental acceptance and cannot be extended to the interview version for children and adolescents. Further research on this topic is desirable, especially regarding the mixed results on the acceptance of children found by Sørensen *et al.* (2007). Second, the outcome variables (overall interviewer and interviewee satisfaction) were constrained to some point because the ratings mostly lay in the higher range of the scales. This may cause a ceiling effect and may influence the generalization of the results of the multilevel analyses. Third, social desirability may play a role even though we tried to secure anonymity. Fourth, interviewees of the two recruitment settings differed in their highest educational level, which may be one reason for the significant difference in their overall satisfaction. Nevertheless, acceptance ratings remained high in both settings and the educational level was no significant predictor in both multilevel models.

Summing up the results, an individual bias of the interviewer on his or her own overall satisfaction, but not on the evaluation of the corresponding interviewee, was found. Furthermore, based on the finding that the acceptance of the interview partner significantly predicted the overall satisfaction with the interview situation, the relationship between both interview partners seems to be important. A professional distance of the interviewer is necessary to not get biased by the interviewee acceptance and to not influence the interviewee acceptance in a negative way by his or her own evaluation. The challenge is to strike a balance between being empathetic and following standardized procedures. Additionally, the interview duration had a significant effect on the interviewer satisfaction. Even though there was only a weak negative correlation between the numbers of conducted interviews and the interview duration, this may reflect that more experience with the interview shortens the interview duration to some extent and has therefore an indirect effect on acceptance.

In conclusion, only factors that can be improved by professional training had an effect on the interviewee and interviewer acceptance. Even though we did not directly test for this, we believe that an intensive training in conducting the interview is an important requirement for achieving the high level of interviewer and interviewee satisfaction that we found in our sample and for assuring reliable and valid diagnoses. In agreement with Birmaher *et al.* (2009, p. 9), we believe that it is a great challenge to train the interviewers to a high standard to enable them "... to make empirically determined 'developmentally appropriate' decisions in coding the KSADS". Furthermore, we suggest giving interviewers the possibility to conduct interviews frequently so that they get more quickly experienced in using the K-SADS-PL. As our results showed, a higher number of conducted interviews may shorten interview durations to some point, which may heighten the interviewer satisfaction even more. Finally, to ensure a consecutive high interview quality after the training process is completed, ongoing supervision is needed. To realize that, we implemented a permanent consulting service into the work schedules of all interviewers. Additionally, we offer regular obligatory supervisions for interviewers in both settings. Through these procedures and despite the high number of interviewers in our study, the inter-rater reliability of diagnoses was on a substantial level and the inter-rater reliability of GAF on a high level.

Although it takes constant effort to keep satisfaction and quality on a high level, we highly recommend using a diagnostic interview. It leads to valid and reliable diagnoses and is well accepted by interviewers and interviewees in different settings.

Acknowledgements

This publication is supported by LIFE – Leipzig Research Centre for Civilization Diseases, Universität Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by means of the

Free State of Saxony within the framework of the excellence initiative.

Declaration of interest statement

The authors have no competing interests.

References

- American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision, DSM-IV-TR, Washington, DC, APA.
- Axelson D., Birmaher B., Zelazny J., Kaufman J., Gill M.K., Brent D. (2009) K-SADS-PL 2009 Working Draft, Pittsburgh, PA, Advanced Center for Intervention and Services Research (ACISR) for Early Onset Mood and Anxiety Disorders, Western Psychiatric Institute and Clinic.
- Beck A.T., Ward C.H., Erbaugh J.K., Mendelson M., Mock J.E. (1962) Reliability of psychiatric diagnoses. A study of consistency of clinical judgments and ratings. *American Journal of Psychotherapy*, **119**(4), 351–357, DOI:10.1007/s10109-003-0097-9
- Birmaher B., Ehmann B., Axelson D.A., Goldstein B.I., Monk K., Kalas C., Kupfer D., Gill M.K., Leibenluft E., Bridge J., Guyer A., Egger H.L., Brent D.A. (2009) Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children – a preliminary psychometric study. *Journal of Psychiatric Research*, **43**(7), 680–686, DOI: 10.1016/j.jpsychires.2008.10.003
- Bruchmüller K., Margraf J., Suppiger A., Schneider S. (2011) Popular or unpopular? Therapists' use of structured interviews and their estimation of patient acceptance. *Behavior Therapy*, **42**, 634–643, DOI: 10.1016/j.beth.2011.02.003
- Delmo C., Weiffenbach O., Gabriel M., Stadler C., Poustka F. (2001) Diagnostisches Interview Kiddie-Sads-Present and Lifetime Version (K-SADS-PL), 5. Auflage der deutschen Forschungsversion, erweitert um ICD-10-Diagnostik [5th edition of the German research version with the addition of ICD-10-diagnosis], Frankfurt, Klinik für Psychiatrie und Psychotherapie des Kindes- und Jugendalters.
- Egger H.L., Ascher B.H., Angold A. (1999) The Preschool Age Psychiatric Assessment. Version 1.1. (Unpublished Interview Schedule), Durham, NC, Center for Epidemiology, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center.
- Ehlert U. (2007) Eine Psychotherapie ist immer nur so gut wie ihre Diagnostik [A psychotherapy is only as good as its diagnosis]. *Verhaltenstherapie*, **17**, 81–82, DOI: 10.1159/000103156
- Endicott J., Spitzer R.L., Fleiss J.L., Cohen J. (1976) The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. [Comparative Study Research Support, US Government, Non-PHS]. *Archives of General Psychiatry*, **33**(6), 766–771, DOI: 10.1001/archpsyc.1976.01770060086012
- Gall S.L., Abbott-Chapman J., Patton G.C., Dwyer T., Venn A. (2010) Intergenerational educational mobility is associated with cardiovascular disease risk behaviours in a cohort of young Australian adults. *The Childhood Determinants of Adult Health (CDAH) Study. BMC Public Health*, **10**, 55, DOI: 10.1186/1471-2458-10-55
- Helzer J.E. (1981) The use of a structured diagnostic interview for routine psychiatric evaluations. *Journal of Nervous and Mental Disease*, **169**(1), 45–49, DOI: 10.1097/00005053-198101000-00006
- Jensen A.L., Weisz J.R. (2002) Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology*, **70**(1), 158–168, DOI: 10.1037/0022-006X.70.1.158
- Joiner J.T.E., Pettit J.W., Walker R.L., Perez M., Cukrowicz K.C. (2005) Evidence-based assessment of depression in adults. *Psychological Assessment*, **17**, 267–277, DOI: 10.1037/1040-3590.17.3.267
- Kaufman J., Birmaher B., Brent B., Rao U., Ryan N. D. (1996) Kiddie-SADS-Present and Lifetime Version (K-SADS-PL), Version 1.0 of October 1996, Pittsburgh, PA: The Department of Psychiatry, University of Pittsburgh School of Medicine.
- Luborsky L. (1962) Clinician's judgments of mental health. *Archives of General Psychiatry*, **7**, 407–417, DOI: 10.1001/archpsyc.1962.01720060019002
- Margraf J., Schneider S., Ehlers A. (1994) Diagnostisches Interview bei psychischen Störungen (DIPS), 2nd edition, Berlin, Springer.
- Marshall R.D., Spitzer R.L., Vaughan S.C., Vaughan R., Mellman L.A., MacKinnon R.A., Roose S.P. (2001) Assessing the subjective experience of being a participant in psychiatric research. *American Journal of Psychiatry*, **158**, 319–321, DOI: 10.1176/appi.ajp.158.2.319
- Pinninti N.R., Madison H., Musser E., Rissmiller D. (2003) MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. *European Psychiatry*, **18**(7), 361–364, DOI: 10.1016/j.eurpsy.2003.03.004
- Quante M., Hesse M., Döhnert M., Fuchs M., Hirsch C., Sergejev E., Casprizg N., Geserick M., Naumann S., Koch C., Sabin M.A., Hiemisch A., Körner A., Kiess W. (2012) For the LIFE child study investigators. *The LIFE child study: a life course approach to disease and health. BMC Public Health*, **12**, 1021, DOI: 10.1186/1471-2458-12-1021
- Raudenbush S.W., Bryk A.S., Cheong A.S., Fai Y.F., Congdon R.T., du Toit M. (2011) HLM 7: Hierarchical Linear and Nonlinear Modelling, Lincolnwood, IL: Scientific Software International.
- Saile H., Weiland-Heil K., Schwenkmezger P. (2000) Can valid diagnoses be reached in initial clinical interviews? Comparison of initial clinical interview, structured interview and symptom checklists. *Zeitschrift für Klinische Psychologie und Psychotherapie*, **29**(3), 214–220.
- Saß H., Wittchen H.-U., Zaudig M., Houben I. (2003) Diagnostisches und Statistisches Manual Psychischer Störungen Textrevision (DSM-IV-TR), Textrevision (DSM-IV-TR), Göttingen, Hogrefe.
- Schneider S., Döpfner M. (2004) Leitlinien zur Diagnostik und Psychotherapie von Angst und Phobischen Störungen im Kindes- und Jugendalter: ein evidenzbasierter Diskussionsvorschlag [Guidelines for diagnosis and psychotherapy of anxiety and phobic disorders in childhood and adolescence: An evidence-based suggestion]. *Kindheit und Entwicklung*, **13**, 80–96, DOI: 10.1026/0942-5403.13.2.80
- Shaffer D., Fisher P., Lucas C., Dulcan M., Schwab-Stone M. (2000) NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV). Description, differences from

- previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, **39**, 28–38, DOI: 10.1097/00004583-200001000-00014
- Shear M.K., Greeno C., Kang J., Ludewig D., Frank E., Swartz H.A., Hanekamp M. (2000) Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry*, **157**(4), 581–587, DOI: 10.1176/appi.ajp.157.4.581
- Silverman W.K., Ollendick T.H. (2005) Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, **34**, 380–411, DOI: 10.1207/s15374424jccp3403_2
- Sørensen M.J., Thomsen P.H., Bilenberg N. (2007) Parent and child acceptability and staff evaluation of K-SADS-PL: a pilot study. *European Child and Adolescent Psychiatry*, **16**(5), 293–297, DOI: 10.1007/s00787-006-0559-y
- Suppiger A., In-Albon T., Hendriksen S., Hermann E., Margraf J., Schneider S. (2009) Acceptance of structured diagnostic interviews for mental disorders in clinical practice and research settings. *Behavior Therapy*, **40**(3), 272–279, DOI: 10.1016/j.beth.2008.07.002
- World Health Organization. (1994) International Statistical Classification of Diseases and Health Related Problems, 10th Revision (ICD-10), Volume III Alphabetical Index, Geneva: World Health Organization.
- Zahner G.E. (1991) The feasibility of conducting structured diagnostic interviews with preadolescents: a community field trial of the DISC. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**(4), 659–668, DOI: 10.1097/00004583-199305000-00023

Supporting information

Additional supporting information maybe found in the online version of this article at the publisher's website.