Vol. 19,　No. 1.　　　　　　　　　　　　　　　　　　　　**March. 2014.**

The KOREAN JOURNAL OF HEALTH PSYCHOLOGY

published by

by Korean Health Psychological Association

Editor : Kyung-Hyun Suh(Sahmyook University)
Associate Editor : Sung-Gun Kang(Kangwonland Addiction Care Center), Young-Gun Go(Korea University),
　　　　　　　　Sun-Jung Kwon(Korea Baptist Theological University), Yong-Hee Kim(Gwangju University),
　　　　　　　　Jhe-Min You(Kangnam University), Min-Kyu Rhee(Gyeongsang National University),
　　　　　　　　In-Hyae Yi(Kangwon National University), Mun-Sun Chang(Kyungpook National University),
　　　　　　　　Sung-Gun Cho(Chungnam National University)

한국건강심리학회 임원진

회 장	현 명 호	(중앙대학교 심리학과)

수석부회장	서 경 현	(삼육대학교 상담심리학과)
국내협력부회장	김 청 송	(경기대학교 청소년학과)
국제협력부회장	서 경 현	(삼육대학교 상담심리학과)

총무이사	조 성 근	(충남대학교)
교육이사	권 선 중	(침례신학대학교 상담심리학과)
홍보이사	최 성 진	(부산메리놀병원)
학술이사	조 현 주	(영남대학교 심리학과)
재무이사	심 은 정	(부산대학교 심리학과)

윤리위원장	박 경	(서울여자대학교 특수치료전문대학원)
편집위원장	서 경 현	(삼육대학교 상담심리학과)
수련위원장	박 준 호	(대전대학교 산업광고심리학과)

선임이사	고 영 건	(고려대학교 심리학과)
	고 진 경	(숭실사이버대학교 상담심리학과)
	김미리혜	(덕성여자대학교 심리학과)
	김 용 희	(광주대학교 심리학과)
	박 경	(서울여자대학교 특수치료전문대학원)
	안귀여루	(강남대학교학교 교육대학원)
	장 문 선	(경북대학교 심리학과)
	정 경 미	(연세대학교 심리학과)
	최 미 례	(육군사관학교 리더십센터)

감 사	이 민 규	(경상대학교 심리학과)
고 문	김 교 헌	(충남대학교 심리학과)

사무국장	주 해 원	(중앙대학교 심리학과)
사무간사	홍 혜 인	(중앙대학교 심리학과)
편집간사	김 선 미	(삼육대학교 상담심리학과)

한국심리학회지
건 강
제 19 권 제 1 호

[Brief Report]

한국건강심리학회

The Validation of the Korean Version of the Knowledge of Evidence-based Services Questionnaire[†]

Chad Ebesutani   Sungwon Choi[‡]

Duksung Women's University

In the present study, we made several necessary improvements of the Knowledge of Evidence-Based Services Questionnaire (KEBSQ)— a measure designed for measure Evidence-Based Practice (EBP) knowledge across various youth problem areas. First, we identified the presence of "bad" items that need to be discarded. Second, we identified the presence of two distinct KEBSQ item types: (1) items associated with "high coverage" and (2) items associated with "low coverage". Results based on exploratory factor analysis, confirmatory factor analysis, and internal consistency estimates clearly showed that these "low coverage" and "high coverage" items have unique properties and should not be combined together to form a total score. Instead, their contrasting properties suggest that "low coverage" and "high coverage" items should comprise of two different versions of the KEBSQ tests. We also significantly reduced the length of KEBSQ by eliminating the "bad" items which create two separate 12-item and 18-item tests to assess knowledge on "low coverage" and "high coverage" treatment practices, respectively. Study implications and additional necessary research efforts are also discussed.

Keywords: Assessment, Psychometrics, Standardization, Evidence-Based Treatments, Child Adolescent treatment

---

Treatment research over the past several decades has led to a body of knowledge that can be used to outline effective mental health treatment practices for youth (Chambless & Hollon, 1998; Lonigan, Elbert, & Johnson, 1998). This evidence-base of effective treatments has been constantly growing ever since and now has broad coverage across various disorders and populations (Chorpita et al., 2002; Chorpita et al., 2011; Hoagwood, Burns, Kiser, Ringeisen, & Schoenwald, 2001; Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Achieving effective dissemination and implementation of new knowledge in applied settings is not easy. In the field of psychology, evidence-based practices (EBPs) have been no exception to this slow adoption process among practicing community clinicians (Daleiden & Chorpita, 2005; Weersing, Weisz, & Donenberg, 2002). It is now clear that systematic efforts are needed to help achieve better dissemination and implementation of EBPs in "real-world" settings (Becker, Nakamura, Young, & Chorpita, 2009; Chorpita & Regan, 2009).

Several efforts have begun relatively recently to enhance the effectiveness of the dissemination of EBP to working clinicians in the field. For example, Chorpita and Daleiden (2009) developed a 'distillation and matching' model which simplifies the ever-growing treatment literature by identifying common practices (known as "Practice Elements") across effective treatment protocols. These practice elements can then be "matched" to clients (based on client characteristics) and can also be used to comprise training models rich in flexibility and trainability, and which can effectively treat multiple co-occurring problems typically seen in community settings (Chorpita, Daleiden, & Weisz, 2005). Due to the likeability (Borntrager et al., 2009) and effectiveness (Weisz et al., 2013) of this modularized approach to mental health treatment delivery, the 'distillation and matching' model and its use of practice elements is leading the field as an effective treatment dissemination model (Chorpita et al., 2013).

As initiatives like this advance forward, one parameter that will be important is the assessment and monitoring of 'knowledge of evidence-based practices.' Theories on information diffusion (cf. Rogers, 2003) have identified the 'knowledge stage' as a crucial component in achieving effective dissemination (in addition to attitudes, adoption decisions, implementation efforts, and sustainability plans). Indeed, without having knowledge of EBPs, it would be difficult to deliver them in practice. Given the importance of being able to assess and monitor 'knowledge of EBPs,' Stumpf and colleagues (2009) developed the Knowledge of Evidence-Based Services Questionnaire (KEBSQ), which is a 40-item self-report measure designed to assess EBP knowledge. The KEBSQ uses a multiple true-false response

format and asks whether various youth treatment techniques are considered "evidence-based" across the problem areas of anxiety, depression, attention/hyperactivity, and disruptive behavior. Since its development, the KEBSQ has been used in various important work, such as in monitoring changes in EBP knowledge following a two-year systems-level quality improvement intervention (Weistetal, 2009). Studies have also been conducted with the KEBSQ among community therapists to understand the relationship between 'knowledge' and 'attitudes' of evidence-based treatments (Nakamura, McMillan, Okamura, & Shimabukuro, 2011). Although the development of the KEBSQ represents an important and much needed step forward in the area of dissemination and implementation science research, the psychometric properties of this measure are still not yet well understood and studies conducted to date on this measure reveal that the psychometric properties and its item composition could be much improved.

There are several reasons why modifications are needed to the KEBSQ. First, no studies to date have sought to identify and eliminate "poor" tests items, such as items that are "too easy" or "too difficult." The KEBSQ is essentially a test (of EBP knowledge), and so, items that are correctly (or incorrectly) endorsed by all(or nearly all) respondents would be "poor" items. Such items would be too easy(or too difficult) for respondents to answer and therefore would provide no information with respect to being able to discriminate between people high and low on the trait. In other words, such items would be unable to discriminate between people high and low on the continuum of EBP knowledge. The KEBSQ is also in need to scale reduction efforts. The KEBSQ has infact been noted to be too long by practitioners (Weistetal., 2009). Removing "poor" items could simultaneously achieve the second important goal of shortening the KEBSQ. This is needed due to its current level of (high) assessment burden (due to the time needed to complete the KEBSQ). Although the KEBSQ includes only 40 items, the instructions of this measure makes the KEBSQ equivalent to having to answer 160 independent Yes/No items. For item #1, as an example, respondents are asked to indicate whether "Exposure" is evidence based for the problem areas of Anxious/Avoidant, Depressed/Withdrawn, Disruptive behavior, and Attention/Hyperactivity problems. There are 40 such items, therefore involving 160 independent choices (i.e., 40 items x 4 yes/no choices = 160 answers).

In addition to needing to get rid of "poor" items and shorten the KEBSQ, a questionable feature found across studies is that the total KEBSQ score has consistently been associated with low reliability (cf. Stumpf et al., 2009, Weistetal., 2009). Specifically, Stumpf and

colleagues (2009) reported unacceptably low test-retest reliability of KEBSQ total score ($r=.56$), followed by Weist and colleagues (2009) reporting "poor" internal consistency of the total score, and Okamura and colleagues (in press) most recently reporting an internal consistency for the total score of .46. Although Stumpf and colleagues (2009) indicated that the KEBSQ items should not necessarily be associated with high internal consistency due to their assertion that the individual KEBSQ test items shouldn't necessarily correlate with each other, it is possible that an appropriate item set has not yet been identified allowing for adequate levels of internal consistency and reliability scores.

A fourth issue with the KEBSQ is that very few factor analytic studies have been conducted to understand its factor structure. The original development study, for example, did not include a factor analysis (Stumpf et al., 2009), and Weist and colleagues (2009) was the first to examine the structure of the KEBSQ, but used principal component analysis (a procedure that is somewhat different than factor analysis; Costello & Osborne, 2005). In their study, however, they used a sample of youth community mental health therapists and found support for "two well-identified components" associated with the KEBSQ items. Notably, however, they reported that approximately a third of the items did not load on either component, speaking to the likelihood of "poor"

items that may need to be removed. The only other factor analytic study conducted to date on the KEBSQ was conducted recently by Okamura and colleagues (in press). In their study, they were the first to employed factor analysis on KEBSQ data and they found that a three factor structure of knowledge fit the data the best—these three factors were labeled "Low Extent and Low Coverage," "High Extent and High Coverage," and "High Extent and Low Coverage." Although this factor analysis conducted by Okamura and colleagues (in press) has moved our thinking forward and has helped us better understand the complicated item properties and factor structure underlying the KEBSQ, it is worth noting that (a) only two of the three factors derived from their factor analysis were associated with adequate reliability; and further, (b) their factor analysis was conducted on the respondent's *raw* responses(as opposed to their scored 'correct' / 'incorrect' answers). Factor analytic studies are therefore needed to be conducted on the (correct/incorrect) scored data. Another feature still not yet explored in factor analytic studies to date is the presence of a single "EBP Knowledge" factor running through the KEBSQ (scored) correct/incorrect data. A single factor structure is most parsimonious and theoretically sufficient to explain variable in (correct/ incorrect) answer responses. Presumably, just as a "Quantitative" factor is likely to be present

and responsible for the variability in the correct/incorrect answers when students complete a math test, it is likely that there is a single "EBP Knowledge" factor that exists and responsible for the variation in correct/incorrect responses when therapists complete the KEBSQ. No studies to date however have examined the factor structure of this single-factor knowledge model based on the correct/incorrect scored KEBSQ data.

Lastly, the KEBSQ is a relatively new measure that is only available in English. There are therefore no measures available to assessment EBP knowledge for service providers who speak other languages. In the present study, we therefore wanted to assess the ability to translate the KEBSQ to into a different language (Korean) for use among a substantially different service provider population other than that used to develop the original KEBSQ in English.

In the present study, we therefore translated the KEBSQ into Korean to evaluate its psychometric properties and refine its item composition and performance. Specifically, we hypothesized that (a) a single-factor structure (an EBP Knowledge factor) would be responsible for the variation in the correct/incorrect (scored) answers, and that (b) when evaluating the factor analytic item properties from this perspective, we would identify "poor" items, as evidenced by low loadings on the single "Knowledge" factor. We also hypothesized that after removing "poor" and problematic items, a one-factor structure would be identified within the scale, and the items of this (reduced) total score would for the first time be associated with adequate reliability and strong fit indices to support scoring and interpreting the KEBSQ total score in future research, trainings, and evaluation initiatives.

Method

Translation

We translated the original English KEBSQ into Korean using the translation procedures recommended by Brislan (1970). First, a bilingual M.A.-level clinical psychology student translated the KEBSQ into Korean; a second bilingual M.A.-level clinical psychology student translated the Korean translations back into English; and discrepancies were then examined by the translation team which consisted of the first author of this article, a clinical psychologist and a graduate student to ensure translation accuracy.

Participants

For the present study, we included participants who were mental health professionals and graduate students in

clinical/health psychology. We included these participants given their background in clinical/health psychology and their potential of having knowledge pertaining to evidence-based treatments. A total of 240 participants filled out the KEBSQ. Among these, 208 participants had no missing data and were included in the present study. The final sample included 190 females (91.8%) and 17 males (8.2%). Ages ranged from 22 to 54 years old (M = 30.0 years; SD = 7.78). All participants were Korean and fluent in Korean. Of the included participants, 82(39.4%) were graduated students in clinical/health psychology programs, 63 (30.3%) were full-time internship trainees in clinical settings, 49 (23.6%) were mental health professionals, and 14 (6.7%) classified themselves as "other" mentalhealth provider.

Measures

Knowledge of Evidence Based Services Questionnaire (KEBSQ; Higa-McMillan, & Chorpita, 2009).

The KEBSQ is a 40-items measure that was designed to assess knowledge of various evidence-based and non-evidence-based techniques for the four youth problem areas of (i) Anxious/Avoidant, (ii) Depressed/Withdrawn, (iii) Disruptive Behavior, and (iv) Attention/Hyperactivity problems. The KEBSQ utilizes a multiple true-false format, somewhat different than typical mental health questionnaires. Each of the 40 items have 4 responses options, each which may be independently endorsed as True or False with respect to whether or not the practice is considered "evidence-based" for the indicated problem area (plus a "None" responses option to indicate that none of the problem areas are considered "evidenced-based for any of the problem areas). To illustrate the scoring procedures, let's use the first item as an example, which asks about the treatment technique of Exposure (i.e., "Introducing the child to a stimulus, either directly or through imagined experience, with the aim of decreasing the child's fear of the object or situation"). Respondents are asked to circle whether Exposure is an evidence-based technique for treating Anxious/Avoidant (A) problems, Depressed/Withdrawan (D) problems, Disruptive behavior (B) problems, and/or (iv) Attention/Hyperactivity (H) problems. The respondent may circle none of these problem areas (and circle "None" if they believe that Exposure is not evidence-based for any of these problem areas); or they may circle one of these problems areas (e.g., if they believe that Exposure is evidence-based only for the area of Anxious/Avoidant problems), or they may circle multiple problem areas (e.g., if the respondent believes that Exposure is an evidence-based treatment for Anxious/Avoidant problems *and* Depressed/Withdrawn problems). For each of

the 40 items, scores therefore range from 0-4, one point given for each of the four answer options correctly endorsed as True or False. The total KEBSQ score therefore ranges from zero to 160 (40 items x 4 independent answer choices), and higher scores indicate more knowledge of EBPs.

The "correct" answers are scored following the procedures outlined in the original development study (Stumpf et al., 2009) and were based off of the most relevant CAMHD (cf. Okamura et al., in press; Stumpf et al., 2009). In the present study, we used the 2009 CAMHD Biennial Report (Chorpita & Daleiden, 2009) since this was the Biennial Report available online most recent in date prior to data collection in 2011. As with other investigations, each of the 40 different practice elements were deemed "evidence-based" for a given problem area if the practice was utilized in 10% or more of all treatment protocols evidencing Level One (Best) or Level Two (Good) support (cf. Okamuraetal., in press; Stumpf et al., 2009).

Although the KEBSQ is a very promising measure to aid in dissemination efforts of EBPs, studies on its psychometric properties have actually demonstrated mixed support. For example, reliability has been consistently low, including low test-retest reliability (Stumpf et al., 2009) and low internal consistency (e.g., Okamura et al., in press; Weisz et al., 2009); on the other hand, the KEBSQ has been found to be able to discriminate between community therapists and clinical psychology graduate students. The KEBSQ was also found to be sensitive to change following a training workshop (Stumpf et al., 2009). However, when used to assess change in EBP knowledge following a two-year system improvement initiative, no changes in knowledge scores were found (Weisz et al., 2009), with the researchers noting that "the psychometric properties of the KEBSQ may have made it difficult to detect effects. The internal consistency of the total score was quite low for a scale of its length, and the median corrected item-total correlation of .25 lay near the .20 threshold where authorities recommend discarding the item, and 17 out of 40 items fell below that threshold."

Procedure

Prior to beginning data collection, the study and all data collection procedures were given approval by the University internal review board. At a Korean psychology conference, we asked participants who were about to partake in an evidence-based practices workshop to participant in the current study. All participants were told that completion of the questionnaires would take approximately 15 minutes and that their participation was completely voluntary. Participants then signed consent forms prior to

filling out the questionnaires. To obtain responses from participants from other mental health profession fields in Korea (e.g., school counselors), we also emailed study information to counselors in Korean elementary, middle, and high school to complete an online survey of the same questionnaire. The 40-item Korean version of the KEBSQ used in the present study may be seen in the Appendix.

Data Analytic Approach

Exploratory Factor Analysis. We conducted exploratory factor analysis (EFA) using Mplus 7.11 (Muthen & Muthen, 2010) to examine the factor structure underlying the Korean KEBSQ data. We conducted EFA on the 40 *scored* KEBSQ items (each with a possible score of 0‐4). We used the Robust Maximum Likelihood (MLR) estimator and Geomin rotation. Given the likely presence of a parsimonious single factor model, we examined the results of a 1-factor exploratory model. To examine support for the 1-factor model, we examined (a) the number of eigenvalues greater than one—and in particular, whether the first eigenvalue was much larger than any other eigenvalues (suggestive of a strong single factor); (b) the strength of fit indices, and (c) item-to-factor loading patterns. We also discarded items that did not have a significantly factor loading on the one-factor model ($p<.01$).

Confirmatory Factor Analysis. We used Mplus 7.11 to conduct confirmatory factor analysis (CFA) to test to the fit of the resulting models based on the EFA described above. We used the Robust Maximum Likelihood (MLR) estimator for the CFA. The following fit indices were evaluated to examine model fit: Root Mean Square Error of Approximation (RMSEA; Steiger, 1990); Comparative Fit Index (CFI; Bentler, 1990), and Tucker-Lewis index (TLI; Tucker & Lewis, 1973). CFI values greater than .90 (Bentler, 1990) and CFI values greater than .95 (Hu & Bentler, 1999) were used as benchmarks for adequate and good model fit, respectively. RMSEA values lower than .08, and lower than .05 were used as cut offs for adequate and good fit, respectively (Browne & Cudeck, 1993). Lastly, weight root-mean-square residual (WRMR) values less than 1.00 were suggested of good model fit (Yu & Muthen, 2002).

Scale Score Reliability. We assessed reliability of the KEBSQ-K scores via examination of alpha coefficients. We used .70 as the cut-off for acceptable reliability (Nunnally, 1978).

Table 1. Factor loadings and standard errors for the 40 KEBSQ-K items.

Item	Content	Factor Loading	S.E.
KEBSQ1	Exposure	−.108	.049
KEBSQ2	Modeling	.643*	.054
KEBSQ3	Relaxation	.588*	.056
KEBSQ4	Therapist Praise/Rewards	.560*	.071
KEBSQ5	Self-Monitoring	.349*	.068
KEBSQ6	Psychoeducation-Child	.244*	.087
KEBSQ7	Activity Scheduling	−.208	.064
KEBSQ8	Skill Building/Behavioral Rehearsal	.113	.065
KEBSQ9	Self-Reward/Self-Praise	.284*	.078
KEBSQ10	Commands/Limit-Setting	.139	.050
KEBSQ11	Psychoeducation-Parent	.671*	.065
KEBSQ12	Response Cost	.439*	.072
KEBSQ13	Tangible Rewards	.369*	.073
KEBSQ14	Parent Praise	.033	.058
KEBSQ15	Parent-Monitoring	−.133	.058
KEBSQ16	Directed Play	−.464*	.063
KEBSQ17	Stimulus/Antecedent Control	.393*	.036
KEBSQ18	Social Skills Training	.774*	.036
KEBSQ19	Family Engagement	−.755*	.068
KEBSQ20	Crisis Management	−.508*	.057
KEBSQ21	Play Therapy	−.559*	.041
KEBSQ22	Supportive Listening	−.721*	.071
KEBSQ23	Parent Coping	−.023	.073
KEBSQ24	Emotional Processing	−.504*	.054
KEBSQ25	Mentoring	−.642*	.051
KEBSQ26	Family Therapy	−.502*	.043
KEBSQ27	Relationship/Rapport Building	.574*	.086
KEBSQ28	Educational Support	.128	.059
KEBSQ29	Maintenance/Relapse Prevention	.408*	.071
KEBSQ30	Peer Modeling/Pairing	−.288*	.058
KEBSQ31	Cognitive/Coping	.467*	.067
KEBSQ32	Natural/Logical Consequences	−.239*	.061
KEBSQ33	Insight Building	−.595*	.081
KEBSQ34	Assertiveness Training	−.110	.058
KEBSQ35	Problem Solving	.552*	.049
KEBSQ36	Time Out	.573*	.078
KEBSQ37	Ignoring or DRO	.324*	.060
KEBSQ38	Communication Skills	.413*	.077
KEBSQ39	Line of Sight Supervision	−.350*	.094
KEBSQ40	Milieu Therapy	−.234	.049

Note. Significant factor loadings at $p < .01$ are indicated with an asterisk.

Results

Exploratory Factor Analysis

Identifying "Poor" Items. The results of the one-factor exploratory factor analysis (EFA) model appear in Table 1. Items with significant factor loadings ($p < .01$) in Table 1 are indicated with an asterisk and were retained for further analysis. Fit indices associated with this 1-factor EFA model were mixed (RMSEA = .054, SRMR = .071, CFI = .745, TLI = .731) suggesting that improvement could be made to the 1-factor model. All items with (low) factor loadings were considered "poor" items and were thus discarded from the scale. An item may be

Table 2. Factor loadings and standard errors for the 30 KEBSQ items.

Item	Content	Factor Loading	S.E.
KEBSQ2	Modeling	.644*	.048
KEBSQ3	Relaxation	.587*	.054
KEBSQ4	Therapist Praise/Rewards	.563*	.056
KEBSQ5	Self-Monitoring	.350*	.070
KEBSQ6	Psychoeducation-Child	.241*	.068
KEBSQ9	Self-Reward/Self-Praise	.284*	.064
KEBSQ11	Psychoeducation-Rarent	.674*	.049
KEBSQ12	Response Cost	.444*	.065
KEBSQ13	Tangible Rewards	.367*	.072
KEBSQ16	Directed Play	-.463*	.058
KEBSQ17	Stimulus/Antecedent Control	.401*	.062
KEBSQ18	Social Skills Training	.776*	.036
KEBSQ19	Family Engagement	-.758*	.035
KEBSQ20	Crisis Management	-.507*	.067
KEBSQ21	Play Therapy	-.551*	.057
KEBSQ22	Supportive Listening	-.722*	.041
KEBSQ24	Emotional Processing	-.497*	.072
KEBSQ25	Mentoring	-.634*	.054
KEBSQ26	Family Therapy	-.501*	.051
KEBSQ27	Relationship/Rapport Building	.576*	.043
KEBSQ29	Maintenance/Relapse Prevention	.407*	.059
KEBSQ30	Peer Modeling/Pairing	-.287*	.071
KEBSQ31	Cognitive/Coping	.471*	.058
KEBSQ32	Natural/Logical Consequences	-.239*	.066
KEBSQ33	Insight Building	-.593*	.061
KEBSQ35	Problem Solving	.554*	.057
KEBSQ36	Time Out	.576*	.049
KEBSQ37	Ignoring or DRO	.325*	.076
KEBSQ38	Communication Skills	.410*	.059
KEBSQ39	Line of Sight Supervision	-.340*	.075

Note. All factor loadings are significant(p<.01) as indicated with an asterisk.

a "poor" item if ,for example, it is too "easy" or too "difficult"—in which cases, such items would not be able to discriminate between individuals high and low on "knowledge." This is akin to including very easy elementary-level math problems(or very difficulty graduate-level math problems) on a highschool math test; nearly all highschool students would get the "very easy" problems correct and the "very difficult" problems incorrect, thereby providing no useful, discriminating information from those items. As an illustrative example from the KEBSQ, the first item ("Exposure") was associated with a non-significant factor loading. This item is likely "too easy," due to most people knowing that Exposure is evidence-based specifically for Anxiety/Avoidance problems (Chorpita & Daleiden, 2009). In fact, the mean score on this ("Exposure") item was very high (M = 3.75, SE =. 57) out of 4.0 possible points —suggesting that most people received all full credit forth is item (i.e., people know that Exposure is "evidence-based" for treating anxiety, but not the other problem areas. This is in contrast to items #2, #3, and #4, for example, which had mean scores of 1.23, 1.34, and 1.44, respectively. This supports the notion that this first Exposure item is "too easy" and should be discarded as a "poor" item. Evaluation of Table 1 thus led to the removal of 10 items due to having non-significant factor loadings on the factor (i.e., items #1, #7, #8, #10,

#14, #15, #23, #28, #34, #40).

We then re-conducted the EFA with the remaining 30 KEBSQ items. Results of this EFA appear in Table 2. As can be seen in Table 2, all items significantly loaded on .Fit indices associated with this 30-item 1-factor EFA model were improved (i.e., RMSEA = .054, SRMR = .063), although revealing that improvement scould still be made to the fit of this 1-factor model (i.e., CFI = .84, TLI = .82). Interestingly, internal consistence of this "1-factor" model was extremely low (alpha = .26).

Another identified issue with these remaining 30 items was the fact that (a) 18 items were associated with positive factor loadings on the single EBP Knowledge factor, while (b) 12 were associated with negative factor loadings on this factor. As a reminder, the factor analysis was conducted not on the participants' raw endorsements(i.e., not on the items they circled), but on the sum of their *scored* responses (ranging from 0-4), indicating the number of correct endorsements for each of the 40 items. Consequently, inclusion of positively and negatively loaded items on a single scale can have adverse consequences—not only on reducing internal consistency, but also on total score scoring and interpretation. To better understand the relation between the items associated with positive and negative factor loadings, we applied two strategies. First, we examined a 'coverage shade map' across the 30

items. Based on the results of this 'coverage shade map,' we then examined a 2-factor model.

Coverage Shade Maps.

To gain a better understanding of the differences between the items associated with positive and negative factor loadings, we looked at the number of problem areas for which each practice element is considered "evidence-based." This concept was recently considered by Okamura and colleagues (in press) and deemed

Table 3. Coverage Shade Map Depicting the Number of Problem Areas for Which Each Item is Considered "Evidence-Based".

Items	Positive Loading Items	Number of problem areas for which each item is considered "evidence-based"				
		None	1	2	3	4
KEBSQ2	Modeling	▓	▓	▓	▓	▓
KEBSQ3	Relaxation	▓	▓	▓	▓	▓
KEBSQ4	Therapist Praise/Rewards	▓	▓	▓	▓	▓
KEBSQ5	Self-Monitoring	▓	▓	▓		
KEBSQ6	Psychoeducation-Child	▓	▓	▓		
KEBSQ9	Self-Reward/Self-Praise	▓	▓	▓		
KEBSQ11	Psychoeducation-Parent	▓	▓	▓		▓
KEBSQ12	Response Cost	▓	▓	▓	▓	
KEBSQ13	Tangible Rewards	▓	▓	▓	▓	
KEBSQ17	Stimulus/Antecedent Control	▓	▓	▓	▓	
KEBSQ18	Social Skills Training	▓	▓	▓	▓	▓
KEBSQ27	Relationship/Rapport Building	▓	▓	▓	▓	
KEBSQ29	Maintenance/Relapse Prevention	▓	▓	▓	▓	
KEBSQ31	Cognitive/Coping	▓	▓	▓	▓	
KEBSQ35	Problem Solving	▓	▓	▓	▓	▓
KEBSQ36	Time Out	▓	▓	▓		
KEBSQ37	Ignoring or DRO	▓	▓	▓		
KEBSQ38	Communication Skills	▓	▓	▓	▓	

Items	Negative Loading Items					
KEBSQ16	Directed Play	▓	▓			
KEBSQ19	Family Engagement	▓				
KEBSQ20	Crisis Management	▓				
KEBSQ21	Play Therapy	▓				
KEBSQ22	Supportive Listening	▓				
KEBSQ24	Emotional Processing	▓				
KEBSQ25	Mentoring	▓				
KEBSQ26	Family Therapy	▓	▓			
KEBSQ30	Peer Modeling/Pairing	▓	▓			
KEBSQ32	Natural/Logical Consequences	▓	▓			
KEBSQ33	Insight Building	▓				
KEBSQ39	Line of Sight Supervision	▓				

relevant to their factor analysis and interpretation of results. They called this concept "coverage," defined as "the extent to which an item on the KEBSQ was considered to be derived from the evidence-based across the four problem areas" (p. 14). In Table 3, we present the 'shade map,' which provides a visual depiction of the "coverage" of the 30 remaining items organized and grouped together according to (a) the positively loaded items(top portion of Table 3) and (b) the negatively (bottom portion of Table 3). For example, the first row in Table 3 is KEBSQ item #2 ("Modeling"), which is an item associated with a positive factor loading. For this item, Modeling is considered an "evidence-based" practiced for all four areas of Anxious/Avoidant, Depressed/Withdrawn, Disruptive behavior, and Attention/Hyperactivity problems (and so the shade map is shaded up through the number four). The main result obtained from the Coverage Shade Map depicted in Table 3 is that the items associated with *positive* factor loadings; and the items associated with negative factor loadings are practices that are evidence-based for very few *(i.e.,* one or no*)* problem areas.

Although the relationship between such item types is complicated, including both types of items in a single scale has the potential to have adverse effects on scoring and interpretation if they are scored together. For example, in the present sample, people on average circled fewer than 2 problem areas (mean = 1.6 problem areas circled; *S.E.* = .26). What this suggests is that, for people low on EBT knowledge, regardless of the problem area(s) they circle as "evidence-based," their endorsements will likely be correct among the items with positive loadings (since positively loaded items are associated with being evidence-based for nearly all problem areas). This thus has the adverse consequence of inflating "knowledge" scores for people "low" on EBT knowledge when they simply guess. This could also explain why internal consistency estimates have been so low in the present sample (alpha = .38), as well as in previous studies (Okamura et al., in press; Stumpf et al., 2009; Weisz et al., 2009).

Confirmatory Factor Analysis

To further understand the association between the items associated with positive and negative loadings, we evaluated a two-factor model using confirmatory factor analysis (CFA), whereby the first factor was comprised of the items associated with positive factor loadings, and the second factor was comprised of the items associated with the negative factor loadings in the initial EFA .Results of this two-factor CFA model revealed decent fit (RMSEA = .049, SRMR = .061, CFI = .861, TLI = .850). All factor loadings were also significant.

More notably, however, the correlation between these two "factors" was very high and negative (r = -.87). These results suggest that these two item types should not be combined into a single factor and scored as is, since both factors provide competing information. As a point of illustration, when computing reliability across all 30 KEBSQ items, internal consistency is extremely low (a = .26). However, when internal consistency for each of the two set of items are examined separately, alpha is much higher (the 18 positively loaded items: a = .82; the 12 negatively loaded items: a = .84). These results provide further support for the idea that there are two distinct types of "knowledge" items (i.e., those that are "evidence based" for multiple problems and those that are "evidence-based" for few to no problem areas) and that these two types of items perform drastically different information from each other in ways that obviate the ability to combine both item types to create a total summed scale score using all 30 items.

Based on these results, it became clear that both item types should not be included in the same KEBSQ test (and should not be combined to create total "Knowledge" score). Rather, the two item types appear to belong to different KEBSQ tests—one related to EBP knowledge for practices associated with "Low EBP Coverage" and one related to EBP knowledge for practices associated with "High EBP Coverage." This is akin to having completely different math tests for high school and college students. This thus led to the creation of two different, shortened and refined final versions of the KEBSQ to be used for different purposes (as described below).

EFA on the 12 "Low-Coverage" KEBSQ-K items

We re-conducted the exploratory factor analysis based on these 12 KEBSQ items associated with "low coverage." Fit indices for this 12-item 1-factor EFA model all met benchmarks for good model fit (RMSEA = .067, SRMR = .054, CFI = .90). Further, as seen in Table 4 , all factor loadings were positive and significant. Eigenvalues were 4.39, 1.12, 1.04, .90, .84, .80, .72, .58, .53, .44, .33, and .32. The first eigenvalue compared to the remaining eivenvalues was much larger, suggestive of the presence of a single factor running through all of these "low coverage" items. As noted above, internal consistency of this 12-item one-factor "low coverage" model was .84 (the first time across all studies to date that the KEBSQ total score was found to be associated with adequate reliability).

EFA on the 18 "High-Coverage" KEBSQ-K items

The EFA results based on the 18 KEBSQ items associated with "high coverage" appear in

Table 4. All fit indices for this 18-item 1-factor EFA model met benchmarks for good model fit (RMSEA = .041, SRMR = .055, CFI = .93). Further, all factor loadings were positive and significant (see Table 4). Eigenvalues were 5.45, 1.34, 1.29, 1.05, 0.96, 0.92, 0.87, 0.82, 0.74, 0.69, 0.65, 0.60, 0.52, 0.52, 0.46, 0.42, 0.38, and 0.32. These eigenvalues also suggest the presence of a single factor running through all 18 items. As noted above, the internal consistency reliability estimate associated with 18-item one-factor "high coverage" model was .82.

Table 4. Factor loadings and standard errors for the 12 "Low-Coverage" and 18 "High-Coverage" KEBSQ-K items.

Item	Content	Factor Loading	S.E.
"Low-Coverage"			
KEBSQ16	Directed Play	.458*	.061
KEBSQ19	Family Engagement	.684*	.053
KEBSQ20	Crisis Management	.554*	.075
KEBSQ21	Play Therapy	.617*	.057
KEBSQ22	Supportive Listening	.698*	.051
KEBSQ24	Emotional Processing	.518*	.085
KEBSQ25	Mentoring	.740*	.045
KEBSQ26	Family Therapy	.503*	.052
KEBSQ30	Peer Modeling/Pairing	.330*	.071
KEBSQ32	Natural/Logical Consequences	.242*	.065
KEBSQ33	Insight Building	.703*	.049
KEBSQ39	Line of Sight Supervision	.431*	.079
"High-Coverage"			
KEBSQ2	Modeling	.666*	.046
KEBSQ3	Relaxation	.593*	.053
KEBSQ4	Therapist Praise/Rewards	.645*	.051
KEBSQ5	Self-Monitoring	.390*	.065
KEBSQ6	Psychoeducation-Child	.230*	.070
KEBSQ9	Self-Reward/Self-Praise	.311*	.065
KEBSQ11	Psychoeducation-Parent	.690*	.049
KEBSQ12	Response Cost	.502*	.068
KEBSQ13	Tangible Rewards	.355*	.074
KEBSQ17	Stimulus/Antecedent Control	.442*	.065
KEBSQ18	Social Skills Training	.765*	.039
KEBSQ27	Relationship/Rapport Building	.546*	.049
KEBSQ29	Maintenance/Relapse Prevention	.397*	.062
KEBSQ31	Cognitive/Coping	.498*	.064
KEBSQ35	Problem Solving	.577*	.054
KEBSQ36	Time Out	.599*	.049
KEBSQ37	Ignoring or DRO	.304*	.077
KEBSQ38	Communication Skills	.394*	.062

Discussion

In the present study, we sought to improve the KEBSQ—a questionnaire designed to measure knowledge of evidence-based youth practices—in multiple ways. First we sought to create a Korean version of this measure to extend the ability to assess knowledge of evidence-based treatments among Korean mental health providers. Although evidence-based treatments have been slow to be adopted (Weersing et al., 2002), they are making their way to the South Korean peninsula, as evidenced by workshops and trainings on evidence-based practices at local Korean conferences increasing in number. Just as the US, South Korea is in need of more effective treatments for youth with mental health problems. For example, suicide rates among individuals in South Korea are marked high and growing (Park & Lester, 2006; Weissman et al., 1999). Effective, evidenced-based services are thus greatly needed to help deal with this and other growing mental health concern in South Korea.

In the current study, we also reduced the administration burden associated with the 40-item KEBSQ measure. This was needed given that the original KEBSQ has been noted to be too long by practitioners (Weist et al., 2009). Completion of the original (40-item) KEBSQ actually involves having to make 160 true/false determinations (i.e., 40 items x 4 independent answer options). The original KEBSQ is thus associated with relatively high administration burden—a characteristic of assessment tools that has been noted to be a significant obstacle in disseminating effective assessment practices (cf. Ebesutani, Bernstein, Chorpita, & Weisz, 2012). The current study is unique in that results warranted the development of two separate versions of the KEBSQ to be used for different purposes, thereby reducing assessment burden associated with each test version. These shortened versions should therefore be able to assess knowledge more efficiently.

Another major outcome of the current study was that this was the first time "poor" items were identified and removed. The removal of poorly performing items is an important step in scale development, and the revised KEBSQ test versions developed in the present study now include only the well-performing items with respect assessing "knowledge of EBPs." As noted above, a related improvement to the KEBSQ accomplished in the present study was the identification of KEBSQ items that fall into two distinct categories associated with different psychometric properties—and thus the development of two separate tests for different assessment purposes. These two item types are (a) KEBSQ items that have several correct answer choices (i.e., having multiple problem

areas for which the item is considered evidence-based—referred to as "high coverage" practices), and (ii) KEBSQ items that have only one or no correct answer choices (i.e., having one or no problems areas for which the item is considered evidence-based—referred to as "low coverage" practices). This study therefore led to the development of two separate, single-factor EBP knowledge measures, each with unique properties and testing purposes: one test targeting knowledge of treatment practices associated with high coverage, and one test targeting knowledge of treatment practices associated with low/no coverage. Each of these one-factor models also fit the data well and were each associated with good reliability (i.e., $a > .80$). Given the development of the two ("high-coverage" and of the KEBSQ test, a remaining question that clinicians and researchers would likely have is "which test should I use?" Ultimately, the answer to this question will be informed by future studies. However, it is likely that the answer will depend on the purpose of the assessment and the target population. For example, in the early phases of assessment monitoring of EBPs in a particular setting, it may be best to use the low/no coverage version n of the KEBSQ. This version can help to simply identify those who know, for example, that certain practices(such as 'play therapy,' 'supportive listening' and 'mentoring') are not considered "evidence-based"

for any of the four problem are as assessed by the KEBSQ(given that among the 12 "low coverage" items, eight items are evidence-based for none of the problem areas). This may be the first domain to assess as individuals learn about EBPs (i.e., that certain practices are not "evidence based"). This "low coverage" KEBSQ version may thus be appropriate for samples with little background or training on EBPs.

On the other hand, when assessing individuals with relatively higher EBP knowledge (e.g., individuals with some learning background on EBPs), the 18-item "high coverage" version of the KEBSQ may be more appropriate than the "low coverage" version. This is because with the "high coverage" items, respondents will (implicitly) be encouraged to make more fine-tune distinctions across the multiple problem areas regarding whether or not each practice is evidence-based. Theoretically and conceptually, this is the next phase in the acquisition and application of EBP knowledge— that is, after simply learning that certain practices are not evidence-based, people then need to learn the more complicated and subtle discrimination rules regarding which treatment practices are evidence-based for which problem areas. These recommendations for KEBSQ version application however are guided largely by theoretical propositions and so it will be important for future studies to test when and for whom the "low coverage" and "high

coverage" versions of the KEBSQ are most appropriate.

Interestingly, in Weist and colleagues' (2009) study, they found support for "two well-identified components (not a single score)" among the KEBSQ items using principal components analysis. Their findings are somewhat consistent with our present findings, in that a single score should not be created based on all KEBSQ items. Based on our results, we specifically make the recommendation that two *separate* test versions be created, consistent with Weistand colleagues' (2009) findings of two separate components. In Weistand colleagues'(2009) study, they also reported that "roughly a third of the items." Although they did not report which items these were that did not load on either component, it is possible that these items were the same "poor" items identified in the present study that needed to be discarded from the measure. It is also likely that the "two components" identified by Weistand colleagues'(2009) refer to the two types of item sets that should comprise their own test version types (i.e., the "low coverage" and "high coverage" items). In their study, Weistand colleagues(2009) however, did not parse out the two components, but summed all items to create a total score(a ssuggested in the original development paper). They reported that the KEBSQ Total score did not show any change at Year 2 following a system

improvement intervention. It is important to note however that this could have been due to either (i) EBP knowledge not actually increasing following the intervention, or (ii) the KEBSQ not consisting of the appropriate items to accurately measure EBP knowledge. Weistand colleagues (2009) themselves asserted that "the psychometric properties of the KEBSQ may have made it difficult to detect effects" following their system-wide quality improvement intervention. The results of the present study suggest that this may be true. Given the present findings (that the "low coverage" and "high coverage" items perform drastically different from each other and should not be combined to form a single total score), it would be important for such analyses to be re-conducted in order to re-evaluate the effectiveness of the intervention while applying this new KEBSQ item scoring framework (i.e., parsing out"low coverage" and "high coverage" items, while discarding "poor" items). Differences may then be found in "knowledge" scores pertaining to the "low coverage" and/or "high coverage" EBP knowledge scores. This illustrates the importance of having a well-performing EBP Knowledge measure(or set of measures) with sound psychometric properties given the implication on evaluations of system-level interventions (cf. Weistetal., 2009).

The present study however was not without

limitations. First, the present study was unable to examine whether the Korean version of the KEBSQ is sensitive to change (such as following a training on evidence-based treatments). Future studies should examine the sensitivity of both KEBSQ Korean versions with respect to sensitivity to change following trainings, as done in the initial development study (Stumpf et al., 2009). Second, studies conducted to date on the psychometric properties of the KEBSQ have been conducted on mental health providers in America; the present study, however, was conducted on Korean individuals in the Korean mental health system. Although we intentionally targeted this sample in order to create the Korean version of the KEBSQ for use among Korean mental health providers, it remains unclear whether the present findings will generalize to samples in other countries. It is thus recommended that, for example, US-based data be (re)analyzed while considering the "low coverage" and "high coverage" scored items and not incorporating them in a total scale score. Reconceptualizing these scoring rules (and item composition) of the KEBSQ can have drastic effects on the total "Knowledge" scores produced, and so more research is needed to examine the generalizability of the present findings. Third, the present sample was comprised of a heterogeneous group of individuals in the Korean mental health field, including

clinical/health psychology graduate students, clinical/health psychology interns working full-time in hospitals and clinics, clinical/health psychology professors and mental health professionals. Although this heterogeneous sample was helpful in developing this initial Korean version of the KEBSQ (due to the heterogeneity in responses it provided for factor analysis), it would be useful for future studies to examine the performance of the Korean KEBSQ in each of these specific subsamples (i.e., clinical/health psychology graduate students, clinical/health interns, and clinical/health psychology mental health professionals). The degree to which the KEBSQ test versions are applicable to each of these domains and participant types will be important for future research to examine. Lastly, although not a limitation of the present study, per se, it is worth making an explicit note about the generalizability of the present findings over time due to the nature of this measure. Specifically, the KEBSQ is essentially a test. And importantly, test items need to be changed, modified, and updated as characteristics of the test takers change over time. For example, the same math test questions should not be given to students as they progress through elementary, middle school, and high school. As individuals progress and increase their knowledge of the target content, the test items should also change accordingly. The same is

true for the KEBSQ. Given the nature of the KEBSQ—particularly the unique item properties associated with the two test versions—it would be ideal for future studies to find ways to develop and administer the KEBSQ via computerized adaptive testing (CAT) approaches that utilize item response theory to identify the most informative item to ask to respondents as they progress through a given test session. This has been recommended elsewhere for other measures (cf. Reise, Morizot, & Hays, 2007), and it would be highly useful for the KEBSQ as well.

Despite these limitations and areas for future research, the present study represents a meaningful step forward in multiple ways related to the KEBSQ. In addition to translating the KEBSQ into Korean for use in the growing Korean mental health field, we have also identified unique item properties associated with "low coverage" and "high coverage" items that justify the creation of two separate versions of the KEBSQ. In addition, each test represents significantly reduced (shortened) versions of the original measure (thereby reducing administration burden) while at the same time yielding a simple one-factor structure that is associated with both strong fit indices and high internal consistency reliability estimates. This thus provides support for scoring total scores for each KEBSQ version. Knowledge of evidence-based treatments is an important

component in the adoption and delivery of effective youth practices. It is thus hoped that the KEBSQ can be a useful tool to aid in assessing and monitoring the learning process among therapists and service providers, leading toward increased knowledge acquisition and effective delivery of evidence-based treatments for youth.

References

Becker, K. D., Nakamura, B. J., Young, J., & Chorpita, B. F. (2009). What better place than here? What better time than now? ABCT's burgeoning role in the dissemination and implementation of evidence-based practices. *The Behavior Therapist, 32*, 89-96.

Bentler, P. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238-246.

Brislin, R. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-159). Newbury Park, CA: Sage.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7-18.

Chorpita, B. F. & Daleiden, E. L. (2009). *2009 Biennial Report: Effective psychosocial interventions for youth with behavioral and emotional needs*. Child and Adolescent

Mental Health Division, Honolulu: Hawaii Department of Health.

Chorpita, B. F. & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77,* 566-579.

Chorpita, B., Daleiden, E., Ebesutani, C., Young, J., Becker, K., Nakamura, B. J., Phillips, L., Ward, A., Lynch, R., Trent, L., Smith, R. L., Okamura, K., & Starace, N. (2011). Evidence Based Treatments for Children and Adolescents: An Updated Review of Indicators of Efficacy and Effectiveness. *Clinical Psychology: Science and Practice, 18*(2), 154-172.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology, 11(3),* 141-156.

Chorpita, B. F. & Regan, J. (2009). Dissemination of effective mental health treatment procedures: Maximizing the return on a significant investment. *Behaviour Research & Therapy, 47,* 990-993.

Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J., Higa-McMillan, C. K, Nakamura, B. J., Austin, A. A., Borntrager, C. F., Ward, A., Wells, K. C. & Gibbons, R. D. (2013). Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting And Clinical Psychology, 81*(6), 999.

Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., McGee, C., Serrano, A., Yates, A., Burns, J. A., & Morelli, P. (2002). Toward large-scale implementation of empirically supported treatments for children: A review and observations by the Hawaii empirical basis to services task force. *Clinical Psychology: Science and Practice, 9*(2), 165-190.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10,* Available Online: http://pareonline.net/getvn.asp?v=10&n=7.

Daleiden, E., Lee, J., & Tolman, R. (2008). *Annual evaluation report: Fiscal year 2004.* Honolulu, HI: Hawaii Department of Health, Child and Adolescent Mental Health Division 2004. Retrieved February 25, 2008, from http://hawaii.gov/health/mental-health/camhd/library/pdf/rpteval/ge/ge/ge011.pdf

Daleiden, E., & Chorpita, B. F. (2005). From Data to Wisdom: Quality Improvement Strategies Supporting Large-scale Implementation of Evidence-Based Services. *Child and Adolescent Psychiatric Clinics of North America, 14,*(2), 329-349,

Ebesutani, C., Bernstein, A., Chorpita, B., & Weisz, J. (2012). A transportable assessment protocol for prescribing youth psychosocial treatments in "real world" settings: Reducing assessment burden via self-report scales. *Psychological Assessment, 24*(1), 141-155.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. A., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health

Research Institute, The National Implementation Research Network, 2005. (FMHI Publication #231).

Hoagwood, K., Burns, B. J., Kiser, L., Ringeisen, H., Schoenwald, S. K. (2001). Evidence-based practice in child and adolescent mental health services. *Psychiatric Services, 52*, 1179-1189.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Lonigan, C., Elbert, J., & Johnson, S. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27,* 138-145.

Muthen, L. K., & Muthen, B. O. (2010). *Mplus User's Guide (4th Ed.).* Los Angeles, CA: Author.

Nakamura, B. J., Higa-McMillan, C. K., Okamura, K. H., & Shimabukuro, S. (2011). Knowledge of and attitudes towards evidence-based practices in community child mental health practitioners. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(4), 287-300.

Okamura, K. H., Nakamura, B. J., Mueller, C., Hayashi, K., & Higa-McMillan, C. K. (in press). An exploratory factor analysis of the knowledge of evidence-based services questionnaire. *Journal of Behavioral Health Services and Research.*

Park, B. C., & Lester, D. (2006). Social integration and suicide in South Korea. *Journal of Crisis Intervention and Suicide Prevention, 27*(1), 48.

Reise, S., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16,* 19-31.

Rogers, E. M. (2003). *Diffusion of innovations (5th ed.).* New York: Free Press.

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research, 25,* 173-180.

Stumpf, R., Higa-McMillan, C., & Chorpita, B. (2009). Implementation of Evidence-Based Services for Youth: Assessing Provider Knowledge. *Behavior Modification, 33,* 48-65.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *Clinical Psychologist, 48,* 3-23.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

Weersing, R. V., Weisz, J. R.., & Donenberg G. R. (2002). Development of the therapy procedures checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child Psychology, 31,* 168-180.

Weissman, M. M., Bland, R. C., Canino, G. J., Greenwald, S., Hwu, H. G., Joyce, P. R., Karam, E. G., Lee, C.-K., Lellouch J., Lepine, J.-P., Newman, S. C., Rubio-stipec, M., Wells, J. E., W ickramaratne, P. J., Wittchen, H.-U., & Yeh, E. K. (1999). Prevalence of suicide ideation and suicide attempts in nine countries. *Psychological medicine, 29*(1), 9-17.

Weist, M., Lever, N., Stephan, S., Youngstrom, E., Moore, E., Harrison, B., Anthony, L., Rogers,

K., Hoagwood, K., Ghunney, A., Lewis, K., & Stiegler, K. (2009). Formative evaluation of a framework for high quality, evidence-based services in school mental health. *School Mental Health, 1*(4), 196-211.

Yu, C., & Muthen, B. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

한국판 근거기반서비스 지식 평가지의 타당화

Chad Ebesutani   최 승 원

덕성여자대학교

본 연구는 다양한 영역 소아문제의 근거기반치료에 대한 지식을 측정하는 근거기반서비스 지식 평가지(KEBSQ)의 개선을 위한 작업이다. 먼저 삭제가 필요한 문항들을 선별하였으며, KEBSQ를 두 유형으로 분리하였다. 이렇게 분리된 두 유형은 (1) '범용' 문항과 (2) '특수 목적용' 문항으로 분류되었다. 탐색적 요인분석, 확인적 요인분석, 내적합치도의 분석결과 이 '범용'과 '특수 목적용' 문항들은 별개의 기능을 하고 있어서 하나의 총점으로 통합하는 것이 적절하지 않음을 확인하였다. 대안으로, 이 두 문항군들은 KEBSQ를 구성하는 별도의 두 검사로 존재하는 것이 타당하다는 결론을 얻게 되었다. 이에 더하여 문항의 역할이 부족한 것으로 분석된 일부 문항들을 제거한 결과 '범용' 문항 18개, '특수 목적용' 문항 12개가 최종적으로 남게 되었다. 마지막으로 본 연구의 시사점과 후속 연구에 대한 제언을 담았다.

주요어: 측정, 심리측정학, 표준화, 근거기반치료, 아동청소년치료

Appendix

<div align="center">

KEBSQ - Original Version
증거기반치료 지식에 대한 설문

</div>

본 설문은 아동 청소년 심리치료에 사용되는 증거기반치료에 대한 여러분의 지식 수준을 알기 위한 목적에서 마련되었습니다. 아동 청소년 심리치료에 사용되는 다음의 치료전략이 불안/회피, 우울/위축, 파괴적(disruptive) 행동, 과잉행동/주의력 문제들 중 어떤 문제에 대한 증거기반치료인지 표시하십시오. 자신이 평소 괜찮다거나 도움이 된다고 생각하는 치료에 답하는 것이 아니라, 연구에 의해 명확하게 검증된 치료에 응답해 주십시오.

증거기반기법이라 생각되는 문제를 <u>모두</u> 고르십시오. A는 불안/회피, D는 우울/위축, B는 파괴적 행동, H는 과잉행동/주의력 문제입니다. 기술된 치료가 증거기반치료가 아니라고 생각되면 N(없음)에 표기하십시오.

A	D	B	H	N	
불안/회피	우울/위축	파괴적 행동	과잉행동/주의력	없음	

예: 클래식 음악을 들려준다.	A	D	B	H	N
1. 사물이나 상황에 대한 두려움 감소를 위해 직접 또는 상상을 통해 아동을 자극에 노출시킨다.	A	D	B	H	N
2. 아동의 모방 및 모방수행 촉진을 위해 바람직한 행동을 모델링을 한다.	A	D	B	H	N
3. 신체적 각성/예민도/반응 감소를 위해 아동에게 근육이완, 호흡연습, 명상과 같은 진정기법을 가르친다.	A	D	B	H	N
4. 아동의 행동 강화를 위해 치료자가 상을 주거나 칭찬을 한다.	A	D	B	H	N
5. 아동에게 자신의 생각, 감정, 행동을 반복적으로 측정하도록 가르친다.	A	D	B	H	N
6. 아동에게 문제의 발생 원인과 치료 근거에 대해 가르친다.	A	D	B	H	N
7. 규칙적으로 즐거운 활동에 참여하도록 아동을 격려한다.	A	D	B	H	N
8. 기술 습득을 목적으로 특정 활동을 한다.	A	D	B	H	N
9. 아동이 바람직한 행동 수행에 대해 스스로 보상하도록 격려한다.	A	D	B	H	N
10. 효과적인 지시와 명령을 할 수 있도록 부모들을 훈련시킨다.	A	D	B	H	N
11. 부모에게 문제의 발생 원인과 치료 근거에 대해 가르친다.	A	D	B	H	N
12. 부정적 행동의 결과로 포인트나 토큰을 차감하는 시스템을 시행한다.	A	D	B	H	N
13. 바람직한 행동 강화를 위해 물질적 보상을 제공하도록 부모를 가르친다.	A	D	B	H	N
14. 바람직한 행동 증진을 위해 칭찬, 격려, 애정 등 사회적 보상을 제공하도록 부모를 가르친다.	A	D	B	H	N
15. 아동의 생각, 행동, 감정을 모니터(체크)할 수 있도록 부모를 가르친다.	A	D	B	H	N
16. 아동-부모의 언어적/비언어적 상호작용 향상을 위해 아동과 특정한 방식으로 놀아주도록 부모를 가르친다.	A	D	B	H	N

17. 문제 행동 감소를 위해 문제 행동의 촉발 요인을 찾고, 이 촉발 요인을 변경시키거나 제거한다.	A	D	B	H	N
18. 대인 관계 기능 향상을 위해 아동에게 사회적 기술을 가르친다.	A	D	B	H	N
19. 가족을 개입시키고 이들의 치료 참여에 대한 긍정적 관심을 고무하기 위한 전략을 사용한다.	A	D	B	H	N
20. 즉각적인 문제 해결과 추적 계획을 통해 위기 상황에 대처한다.	A	D	B	H	N
21. 주요 치료기법으로 놀이치료를 제공한다.	A	D	B	H	N
22. 경청하고 반영(reflective discussion)하여 따뜻함, 공감, 긍정적 존중을 보여준다.	A	D	B	H	N
23. 부모에게 스트레스적 상황에 대한 대처 전략을 가르친다.	A	D	B	H	N
24. 기존 기억을 새롭고 양립할 수 없는 정보로 대체하여 아동의 정서 처리를 돕는다.	A	D	B	H	N
25. 아동에게 긍정적 롤모델이 될 수 있는 멘토를 제공한다.	A	D	B	H	N
26. 가족 구성원들의 대인관계 및 상호작용 개선을 목적으로 가족치료를 제공한다.	A	D	B	H	N
27. 치료자와 아동 사이의 라포형성을 위한 전략을 시행한다.	A	D	B	H	N
28. 숙제 또는 학습기술과 같은 특정 학업 문제를 다루기 위해 아동에게 보조교사나 튜터를 제공한다.	A	D	B	H	N
29. 치료 이득의 손실 최소화를 위해, 이미 개발된 기술을 강화시키고 추후 생길 수 있는 문제를 예상해본다.	A	D	B	H	N
30. 상호 간의 학습 또는 기술 연습을 돕기 위해, 아동을 또래와 매치한다.	A	D	B	H	N
31. 아동의 생각의 정확성을 평가하거나 해석을 바꾸기 위해 고안된 전략을 사용한다.	A	D	B	H	N
32. 바람직하지 않은 행동에 뒤따르는 나쁜 결과를 아동이 경험할 기회를 주도록 부모를 교육한다.	A	D	B	H	N
33. 아동이 통찰력과 더 많은 자기 이해를 개발하도록 가르친다.	A	D	B	H	N
34. 아동에게 자기주장 기술을 가르치고 다른 이에게 자신의 주장을 할 수 있게 연습시킨다.	A	D	B	H	N
35. 아동에게 문제 찾기, 다수의 해결책 만들기, 최고의 대안책 선택과 같이 단계로 나누어 문제를 해결하도록 가르친다.	A	D	B	H	N
36. 바람직하지 않은 행동에 대한 결과로 타임아웃 기법을 사용한다.	A	D	B	H	N
37. 경미한 수준의 부적절한 행동은 선택적으로 무시하고 대안적 행동에 주의를 기울이도록 부모를 가르친다.	A	D	B	H	N
38. 부모와 아동의 의사소통 개선을 위해 적극적인 경청과 나-전달법("I" statement)과 같은 특정 기법을 가르친다.	A	D	B	H	N
39. 아동의 안전과 적절한 행동 유도를 위해 부모가 아동을 자신의 시야 안에 두게 한다.	A	D	B	H	N
40. 치료의 일환으로 환경을 이용하기 위해, 거주형 치료 시설에서 치료한다.	A	D	B	H	N

Appendix

KEBSQ - "Low Coverage" Version
증거기반치료 지식에 대한 설문

본 설문은 아동 청소년 심리치료에 사용되는 증거기반치료에 대한 여러분의 지식 수준을 알기 위한 목적에서 마련되었습니다. 아동 청소년 심리치료에 사용되는 다음의 치료전략이 불안/회피, 우울/위축, 파괴적(disruptive) 행동, 과잉행동/주의력 문제들 중 어떤 문제에 대한 증거기반치료인지 표시하십시오. 자신이 평소 괜찮다거나 도움이 된다고 생각하는 치료에 답하는 것이 아니라, 연구에 의해 명확하게 검증된 치료에 응답해 주십시오. 증거기반기법이라 생각되는 문제를 <u>모두</u> 고르십시오. A는 불안/회피, D는 우울/위축, B는 파괴적 행동, H는 과잉행동/주의력 문제입니다. 기술된 치료가 증거기반 치료가 아니라고 생각되면 N(없음)에 표기하십시오.

A	D	B	H	N	
불안/회피	우울/위축	파괴적 행동	과잉행동/주의력	없음	

	A	D	B	H	N
1. 아동–부모의 언어적/비언어적 상호작용 향상을 위해 아동과 특정한 방식으로 놀아주도록 부모를 가르친다.	A	D	B	H	N
2. 가족을 개입시키고 이들의 치료 참여에 대한 긍정적 관심을 고무하기 위한 전략을 사용한다.	A	D	B	H	N
3. 즉각적인 문제 해결과 추적 계획을 통해 위기 상황에 대처한다.	A	D	B	H	N
4. 주요 치료기법으로 놀이치료를 제공한다.	A	D	B	H	N
5. 경청하고 반영(reflective discussion)하여 따뜻함, 공감, 긍정적 존중을 보여준다.	A	D	B	H	N
6. 기존 기억을 새롭고 양립할 수 없는 정보로 대체하여 아동의 정서 처리를 돕는다.	A	D	B	H	N
7. 아동에게 긍정적 롤모델이 될 수 있는 멘토를 제공한다.	A	D	B	H	N
8. 가족 구성원들의 대인관계 및 상호작용 개선을 목적으로 가족치료를 제공한다.	A	D	B	H	N
9. 상호 간의 학습 또는 기술연습을 돕기 위해, 아동을 또래와 매치한다.	A	D	B	H	N
10. 바람직하지 않은 행동에 뒤따르는 나쁜 결과를 아동이 경험할 기회를 주도록 부모를 교육한다.	A	D	B	H	N
11. 아동이 통찰력과 더 많은 자기 이해를 개발하도록 가르친다.	A	D	B	H	N
12. 아동의 안전과 적절한 행동 유도를 위해 부모가 아동을 자신의 시야 안에 두게 한다.	A	D	B	H	N

Appendix

<div align="center">

KEBSQ - "High Coverage" Version
증거기반치료 지식에 대한 설문

</div>

본 설문은 아동 청소년 심리치료에 사용되는 증거기반치료에 대한 여러분의 지식 수준을 알기 위한 목적에서 마련되었습니다. 아동 청소년 심리치료에 사용되는 다음의 치료전략이 불안/회피, 우울/위축, 파괴적(disruptive) 행동, 과잉행동/주의력 문제들 중 어떤 문제에 대한 증거기반치료인지 표시하십시오. 자신이 평소 괜찮다거나 도움이 된다고 생각하는 치료에 답하는 것이 아니라, 연구에 의해 명확하게 검증된 치료에 응답해 주십시오. 증거기반기법이라 생각되는 문제를 <u>모두</u> 고르십시오. A는 불안/회피, D는 우울/위축, B는 파괴적 행동, H는 과잉행동/주의력 문제입니다. 기술된 치료가 증거기반치료가 아니라고 생각되면 N(없음)에 표기하십시오.

A	D	B	H	N			
불안/회피	우울/위축	파괴적 행동	과잉행동/주의력	없음			

2. 아동의 모방 및 모방수행 촉진을 위해 바람직한 행동을 모델링을 한다.	A	D	B	H	N
3. 신체적 각성/예민도/반응 감소를 위해 아동에게 근육이완, 호흡연습, 명상과 같은 진정기법을 가르친다.	A	D	B	H	N
4. 아동의 행동 강화를 위해 치료자가 상을 주거나 칭찬을 한다.	A	D	B	H	N
5. 아동에게 자신의 생각, 감정, 행동을 반복적으로 측정하도록 가르친다.	A	D	B	H	N
6. 아동에게 문제의 발생 원인과 치료 근거에 대해 가르친다.	A	D	B	H	N
9. 아동이 바람직한 행동 수행에 대해 스스로 보상하도록 격려한다.	A	D	B	H	N
11. 부모에게 문제의 발생 원인과 치료 근거에 대해 가르친다.	A	D	B	H	N
12. 부정적 행동의 결과로 포인트나 토큰을 차감하는 시스템을 시행한다.	A	D	B	H	N
13. 바람직한 행동 강화를 위해 물질적 보상을 제공하도록 부모를 가르친다.	A	D	B	H	N
17. 문제 행동 감소를 위해 문제 행동의 촉발 요인을 찾고, 이 촉발 요인을 변경시키거나 제거한다.	A	D	B	H	N
18. 대인 관계 기능 향상을 위해 아동에게 사회적 기술을 가르친다.	A	D	B	H	N
27. 치료자와 아동 사이의 라포형성을 위한 전략을 시행한다.	A	D	B	H	N
29. 치료 이득의 손실 최소화를 위해, 이미 개발된 기술을 강화시키고 추후 생길 수 있는 문제를 예상해본다.	A	D	B	H	N
31. 아동의 생각의 정확성을 평가하거나 해석을 바꾸기 위해 고안된 전략을 사용한다.	A	D	B	H	N
35. 아동에게 문제 찾기, 다수의 해결책 만들기, 최고의 대안책 선택과 같이 단계로 나누어 문제를 해결하도록 가르친다.	A	D	B	H	N
36. 바람직하지 않은 행동에 대한 결과로 타임아웃 기법을 사용한다.	A	D	B	H	N
37. 경미한 수준의 부적절한 행동은 선택적으로 무시하고 대안적 행동에 주의를 기울이도록 부모를 가르친다.	A	D	B	H	N
38. 부모와 아동의 의사소통 개선을 위해 적극적인 경청과 나-전달법("I" statement)과 같은 특정 기법을 가르친다.	A	D	B	H	N