

A Training-Model Scale's Validity and Reliability Coefficients: Expert Evaluation in Indonesian Professional Psychology Programs

Anrilia E. M. Ningdyah^{1*}, Kenneth Mark Greenwood², and Garry Kidd³

1. James Cook University, College of Healthcare Sciences, Department of Psychology James Cook Drive, Townsville, QLD 4811, Australia

2. University of the Sunshine Coast, School of Social Sciences, Maroochydore DC, QLD 4558, Australia

3. James Cook University, College of Healthcare Sciences, Department of Psychology Cairns Campus, 14-88 McGregor Road, Smithfield, QLD 4878, Australia

*E-mail: anrilia.ansyorideas@my.jcu.edu.au

Abstract

Very little information has been available on training models in professional psychology programs in Indonesia, despite the Indonesian National Accreditation Body recommending that scientist-practitioner models be applied in the education of psychologists. By contrast, research abounds on such training models in Western countries. This discrepancy raises the importance of developing a measurement tool appropriate for assessing training models in Indonesian professional psychology programs. This article describes the process of testing the validity and reliability of such a training model measuring tool in the Indonesian context. The authors used the expert evaluation method and the Aiken formula to calculate a coefficient of content validity and item's internal consistency reliability. This process formed a training model scale comprising 77 items with satisfactory validity and reliability indexes for measuring Indonesian professional psychology program training models.

Pengujian Validitas Isi dan Reliabilitas Alat Ukur Model Pendidikan pada Program Profesi Psikologi di Indonesia: Penggunaan Metode Evaluasi oleh Panel Ahli

Abstrak

Informasi mengenai model pelatihan dalam program profesi psikologi di Indonesia masih terbatas meskipun Badan Akreditasi Nasional Indonesia menetapkan penggunaan model "ilmuwan-praktisi" (*scientist-practitioner model*) untuk diterapkan di pendidikan profesi psikolog. Padahal, penelitian mengenai model tersebut telah banyak dilakukan di negara-negara Barat. Ketimpangan tersebut meningkatkan urgensi pengembangan alat ukur yang tepat untuk menilai model-model pendidikan yang digunakan dalam program profesi psikologi di Indonesia. Artikel ini menggambarkan proses pengujian validitas dan reliabilitas alat ukur karakteristik model-model pendidikan tersebut dengan nuansa Indonesia. Para peneliti menggunakan metode evaluasi oleh ahli dan formula Aiken untuk menghitung koefisien validitas konten dan reliabilitas konsistensi internal item-item yang ada. Hasil dari proses tersebut adalah terkonstruksinya skala model pendidikan dengan 77 item yang memiliki indeks validitas dan reliabilitas yang memuaskan sehingga dapat digunakan untuk mengukur model pendidikan pada program profesi psikologi di Indonesia.

Keywords: content validity, homogeneity reliability, professional psychology, training model

Citation:

Ningdyah, A. E. M., Greenwood, K. M., & Kidd, G. (2018). A training-model scale's validity and reliability coefficients: Expert evaluation in Indonesian professional psychology programs. *Makara Human Behavior Studies in Asia*, 22(1), 56-66. doi: 10.7454/hubs.asia.2190318

1. Introduction

In a professional psychology program the training model is an important educational component for aspiring psy-

chologists playing a significant role in determining educational direction (Horn et al., 2007) by guiding formation of program objectives and determining learning experiences to achieve them. The training model colors

educational activities: research (e.g., Cherry, Messenger, & Jacoby, 2000) shows that different training models produce graduates with different characteristics.

Academic discussion on training models in professional psychology education dates from 1948, at which time American scientists and practitioners in psychology formulated a training model to provide standards for implementation of education for professional psychologists (Baker & Benjamin, 2000; Cautin & Baker, 2014). The first such training model—the scientist-practitioner model—was developed in that era, and many professional psychology schools worldwide have adopted it (Baker & Benjamin, 2000; Belar & Perry, 1992; Bell & Hausman, 2014; Horn et al., 2007). The scientist-practitioner model balances emphasis on practice and science/research components in educating future psychologists, endorsement of educative integration activities that connect science and practice (Belar & Perry, 1992).

Dissatisfaction with the scientist-practitioner training model led to formation of others, the most common being the practitioner model, which emerged in 1973 (Korman, 1974). Its derivatives include the practitioner-scholar (Bell & Hausman, 2014; Ellis, 1992; Rodolfa, Kaslow, Stewart, Keilin, & Baker, 2005), the local-clinical scientist models (Stricker, 1997; Stricker & Trierweiler, 1995), and the Clinical-Science model created in the 1990s (McFall, 1991, 2006). The practitioner model emphasizes the practice component (Korman, 1974). Deriving from the practitioner model, the practitioner-scholar model gives greater attention to the science/research aspect than its prototype, while still emphasizing the practice component and practical application of scientific knowledge over science/research (Ellis, 1992; Rodolfa, et al., 2005). The local-clinical scientist model, in addition to emphasizing psychological practice, places greater importance on delivering psychological services according to clients' specific needs (Bell & Hausman, 2014). By contrast, the Clinical-Science model emphasizes science and research components over practice (McFall, 2006).

At the end of the 1980s, training programs placed importance on considering graduates' competency in conducting psychological practice, rather than merely emphasizing learning materials and students' practice hours during their professional education. The training models above focused more on which learning content should be emphasized in professional programs (McEvoy et al., 2005), thus relating to the classical debate over whether practice or science/research should be deemed more important in educating future psychologists. Thus, these previous training models are classified as content-based models. Recent discussion has shifted to the set of competencies graduates must attain on completing a professional program. In turn, target competencies are now the main driver determining necessary learning experiences in professional education. The determination

of target competencies and indicators of attainment of those competencies, along with competency measuring tools, are central to competency-based models in professional psychology education. The new competency culture continues to grow and develop to the point that some scholars (e.g., Rodolfa et al., 2014; Rodolfa, et al., 2005) and professional organizations (British Psychological Society, 2015; Canadian Psychological Association, 2001, 2011; National Council of Schools and Programs of Professional Psychology, 2014) have succeeded in formulating a set of target competencies accompanied by behavioral indicators for each competency, and supplemented by formulation of measuring tools and methods to determine each learner's level of competence (Fouad et al., 2009). For a more complete discussion on training models in professional psychology education and their respective characteristics, please refer to Ningdyah, Greenwood, Kidd, Helmes, and Thompson (2016).

In contrast to Western countries, professional psychology program providers and educators in Indonesia are relatively unfamiliar with the notion of training models, information on which is only rarely available. Of the 19 professional psychology programs in Indonesia, only one explicitly mentions its training model (Universitas Surabaya, 2015). General information about learning content is available in other programs, but the specific training model applied in these programs is not available to the public, either as of brochures or on official websites. Statements on specific training models are useful in providing an overview not only of the learning content a program provides and emphasizes in a program, but also on the nature of internship as an important component in professional psychology education (Sullivan & Conoley, 2001). In its current guidelines on professional psychology programs' accreditation mechanism, the Indonesian National Accreditation Board (BAN-PT, 2013a) as the only institution implementing accreditation for these programs, has stated that the scientist-practitioner model should be used in Indonesian professional psychology programs. Accreditation instruments designed and used by BAN-PT (2013b) were also structured along the lines of the scientist-practitioner model.

Besides lack of information on training models in Indonesian programs, research profiling Indonesian professional psychology programs is scarce, if not absent. Such research is abundant in Western countries, including on professional programs' basic profiles (e.g., Pachana, O'Donovan, & Helmes, 2006), characteristics of graduates (e.g., Cherry et al., 2000), and students' views of their experience during professional education (e.g., Merlo, Collins, & Bernstein, 2008). Expert evaluation of a training models instrument as discussed in this article is part of a comprehensive study that attempts to overview the basic profile and curricula characteristics of Indonesian professional psychology programs.

In researching training models and other related elements of professional psychology programs, previous researchers (e.g., Merlo et al., 2008; Nixon, 1994; Pachana et al., 2006; Rodolfa, et al., 2005) have mostly used cross-sectional surveys with questionnaires designed specifically to answer research questions. In those studies, questions to detect training models applied in target programs were formulated as closed-ended or mixed-question format, thus providing respondents with multiple choice options and an opportunity to add responses other than those already provided (see for example, Rodolfa, et al., 2005). In the Indonesian context, this questioning technique may not be appropriate due to the prevailing lack of familiarity with the training-model concept. Some fear that study results might be hampered by respondents being unable to identify accurately the type of training model actually used in their program. Leong and Zachar (1991) used another item-generation approach in their study, which constructed items based on characteristics of a specific training model. This approach was deemed more appropriate for this research, since it overcomes the weakness of direct questioning techniques.

Results of a literature review on the main training models in professional psychology education (Ningdyah et al., 2016) formed the basis of item development in the current investigation's training-model scale. Items in the training-model questionnaire were classified into six groups corresponding with the six dominant training models identified in the literature: 1) scientist-practitioner; 2) practitioner; 3) practitioner-scholar; 4) local-clinical scientist; 5) clinical-science; and, 6) competency based. An evaluation to assess the validity and reliability of the training-model scale that is this article's content was conducted to test the scale's effectiveness in the Indonesian context.

A content validity test was applied in evaluating the training-model scale, to assess how well the measuring instrument represents the relevant content areas (Haynes, Richard, & Kubany, 1995; Rosnow & Rosenthal, 2002) and to ensure that material not relevant to measurement purposes was not included (Azwar, 2012). Reliability testing of the training model scale drawing on the concept of internal-consistency reliability, was intended to ascertain that items measuring the same general construct actually produce similar scores. Moreover, the measuring tool's face validity was assessed to ensure that the training-model scale's format was such that respondents would be motivated to participate (Rosnow & Rosenthal, 2002) and that words, sentences, and terms were appropriate for the Indonesian context.

Availability of a valid reliable training-model scale adapted specifically for use in Indonesia is indispensable, considering the absence of such an instrument and the lack of relevant studies in professional psychology education there. The training-model scale developed from this study's results is assumed beneficial for identifying

characteristics of training model(s) applied in Indonesian professional psychology programs. Description of a program's training-model(s) profile obtained from the scale's use can serve as input for development of educational processes in a particular professional program. This applies especially to implementation of science and practice integration, an important requirement for application of the scientist-practitioner model- the model of professional psychology education that the Indonesian government requires.

2. Methods

The expert evaluation method was used to determine the training-model items' content validity and internal-consistency reliability. Experts in professional psychology education in Indonesia were invited to assess each item's degree of relevance to the training-model component it was intended to represent.

Two popular methods can be used to calculate content validity based on expert judgment. Lawshe (1975) first proposed the content-validity ratio for quantifying the rating of experts' judgment on items. This content-validity ratio (CVR) yields a value from -1.00 to $+1.00$. Lawshe provides a table of significance that contains critical CVR values to determine the degree of content validity of the obtained CVR in accordance with the number of experts involved and with a degree of significance at the 0.05 and 0.01 levels.

Another approach to calculating items content-validity value based on expert evaluation was proposed by Aiken (Aiken, 1980, 1985). Similar to Lawshe's, Aiken's content-validity coefficient is calculated based on experts assessment of item's relevance in measuring the intended construct, with rating categories arranged in a Likert-scale format.

The Lawshe (1975) formula tends to be difficult to implement because it requires a large number of experts to be involved in order for an item to be deemed significant at an acceptable CVR value. The fewer the appraisers, the greater the CVR required. As an illustration, in the table of significance provided by Lawshe (1975, p. 568), if there are only seven experts, an item requires a CVR of 0.99 to be significant and have adequate content validity at a significance level of 0.05. Employing a large number of experts to ensure the critical value demanded is not too high is unrealistic in this study, since the availability of experts in professional psychology education is limited. Azwar (2012) proposes interpretation of the CVR value within its relative range, that spanning from -1 to $+1$. Items with negative values are considered to have very low content validity, so they need to be removed from the measuring instrument, while items with positive CVR value are considered to have content validity at a certain level. However, such interpretation

is susceptible to subjectivity and thus risks generating subjective results and lowering interpretation standards consistency (Yu, 1993, in Yang, 2011). Accordingly, the content validity formula proposed by (Aiken, 1980, 1985) was used in this study.

To assess expert degree's of consistency in evaluating items, Aiken's (1985) statistical formula, the homogeneity coefficient, was also used in this study. The homogeneity coefficient (H) serves as an internal-consistency reliability coefficient for rating data (Aiken, 1985).

Participants and procedure. Approval from the university's human research ethics committee was obtained before data collection commenced. Experts were recommended by the Indonesian Psychological Association (HIMPSI), the sole professional psychology organization in Indonesia involved in of accreditation of professional programs. Selection of experts was based on the following criteria: 1) they were HIMPSI members who have been actively involved in preparation of accreditation instruments for professional psychology programs; or, 2) they were HIMPSI members who have been involved in accreditation of professional programs with the national accreditation body; or, 3) they were academics with current or past involvement in management of professional psychology programs, but not serving as program directors at the time of data collection. HIMPSI recommended eight experts. Via email, the researcher invited each expert to participate in the training-model evaluation study, sending details of research objectives and statements about the importance of their participation in developing a measuring tool to identify training models in Indonesian professional psychology programs. Six experts responded to this invitation, expressed their willingness to participate in the research, and signed informed consent forms. They were then emailed the questionnaire. By the end of the data collection period, five questionnaires had been returned and all were considered valid for further analysis.

Measures. A specific questionnaire was developed for this study, the Expert Evaluation Form for the Training-Models Scale. This measuring tool consists of six item clusters arranged according to the six previously specified types of training models identified in the literature review (Ningdyah et al., 2016), abbreviated as follows: 1) the scientist practitioner model (SP); 2) the practitioner model (P); 3) the practitioner-scholar model (PS); 4) the local-clinical scientist (LCS) model; 5) the Clinical-Science model (CS); and, 6) the competency-based model (CB). To assess items' content validity, respondents were asked to determine the extent to which each item was relevant to the training model it represented, with ratings from 1 (completely irrelevant) to 5 (extremely relevant). In addition, experts were invited to write specific comments regarding items or the measuring tool, to improve item quality in particular and the measuring tool as a whole, in yielding responses.

Data analysis. Aiken's (1980, 1985) content validity index (V coefficient) was calculated for all items on each training model component by applying the following formula:

$$V = S/[n(c-1)] \text{ (Aiken, 1985, p. 133)}$$

where S represents the sum of the absolute values of the difference of each rating by the appraiser, with n the number of raters and c the number of rating categories.

V coefficient ranges from 0 to 1: The greater the V, the higher an item's content validity. Aiken (1985) provides a table of significance for determining the value of V which can be considered significant closest to .05 and .01 for a given number of raters and of rating categories.

To determine the extent to which experts agreed regarding an item's relevance, Aiken's homogeneity coefficient (H) for each item was calculated with the following formula (1985):

$$H = 1 - 4S/[(c-1)(n2-j)] \text{ (Aiken, 1985, p. 140)}$$

where S represents the sum of the absolute values of the difference of each rating by the appraiser, n is the number of raters, $j = 0$ if n is an even number and $j = 1$ when n is an odd number.

The Aiken's H coefficient presented above quantifies the degree of expert's consistency of in assessing an item. The H value ranges from 0 to 1. Similar to the value of V, Aiken also provides a table of statistical significance to determine the critical value of the H coefficient that is considered significant for a given number of rating categories and of raters, at a significance level closest to 0.05 and 0.01.

3. Results

Table 1 in appendix presents a summary of statistical calculations of experts' evaluation results including the mean, standard deviation, the Aiken's Content-Validity Coefficients (V), and the Homogeneity Coefficients (H).

The value of V coefficients in the SP component range from 0.65 to 1.00. In the P component, coefficient Vs range from 0.85 to 1.00. In the PS group, V coefficients range from 0.95 to 1.00. V coefficients in the LCS group range from 0.90 to 1.00. For the CS component, V coefficients were in the range of 0.85 to 1.00. Lastly, V coefficients for the CB models component range from 0.75 to 1.00.

Checking the Aiken's significance table (1985, p. 134) for critical V value shows that for five experts and a five- category evaluation rating, the validity coefficient (V value) must be greater than 0.80 to have a sufficient

content validity ($p = 0.05$). Of the 195 items, five had smaller V values than this critical value; thus, these items were eliminated from the measuring instrument.

For homogeneity reliability, H coefficients within the SP group range from 0.17 and 1.00. H coefficients for the P component range from 0.58 to 1.00. In the PS group, values of Hs range from 0.83 to 1.00. In the LCS group, H coefficients range from 0.75 to 1.00. H coefficients for the CS group range from 0.58 to 1.00, and for the CB group H values range from 0.17 to 1.00.

Checking against Aiken's significance table shows that for five experts and a five category evaluation rating, the homogeneity coefficient (H value) of an item must be greater than 0.75 to be deemed significant ($p = 0.05$).

From the 195 items, 28 had lower H values than the critical H demanded (these 28 items included the five items with low content validity value mentioned above). All items with these low V and H coefficients were eliminated from the training-model scale.

Experts provided useful comments suggesting how the measuring instruments could be improved. Their comments on particular item(s) or the measuring instrument as a whole are presented in Table 2 (appendix).

Experts' specific, item-related comments to suggest using more appropriate words and expressions within the Indonesian context. Items' wording was improved in direct response to experts' comments and advice. Furthermore, as in Table 2 (appendix), experts' general comments primarily related to the scale's length, repetition of some items, and of some overlapping items, so, experts suggested shortening the scale by removing similar and overlapping items. Subsequent comprehensive analysis included re-examining each item's content and the training-model domain of a number of repetitive items. Accordingly, several items were combined, repetitive items were deleted, and the use of words and sentences was again reviewed to produce easier-to-understand items. The final result was development of 77 items, with the following psychometric properties: Content-validity coefficients range 0.85 to 1.00; internal reliability coefficients range from 0.75 to 1.00.

4. Discussion

A measuring tool for training models in professional psychology programs is needed in Indonesia where little, if any, research on these programs has been undertaken. The expert evaluation study discussed in this article was part of a project to develop a valid, reliable training-model scale for the Indonesian context.

The scale's original design included 195 items. Analysis of expert opinion calculations using Aiken's V and H

coefficients showed that 28 items had lower V and H coefficients than required, and these were deleted. Typically, deleted items were too general or vague and deemed unsuitable by experts in distinguishing different types of training models. Some were also affirmative items that referred to common characteristics or facts initially thought applicable to Indonesian professional psychology programs. For example, basic psychological practice content is taught in the program's early years (SP component, number 59).

Our study's results demonstrate that use of expert judgment methods, including careful selection of experienced experts, is a very beneficial part of item selection. Due to their experience in professional psychology program accreditation and management, experts could provide valuable input when evaluating items in the training-model scale. The overall judgment process identified items that could have lowered the scale's efficacy in discriminating among training models, so these items were eliminated.

Expert respondents' comments on specific items included suggestions on word selection, so their more precise wording was adopted. Rigorous effort was also directed towards modifying or eliminating items they regarded as repetitive and overlapping. The re-examination procedure also included reviewing each training model's characteristics, particularly classification in content-based models.

Content-based models emphasize programs' educational content and classify models accordingly, in terms of the practice component, the science/research component, or both. A different emphasis on learning content leads to a different training model. As previously mentioned, for example, emphasis on the science/research component is one of the CS model's main characteristics. However, thorough examination of training models' characteristics shows that content-based training models are not mutually exclusive; emphasizing one content component does not necessarily completely eliminate other content components. In the practitioner model, for example, the dominant emphasis on the psychological practice component does not automatically preclude delivery of the science component. Korman (1974), a supporter of this model, states that, despite the fact that the practitioner model emphasizes the practice component and the delivery of psychological service, educational experiences were delivered to students "...without abandoning comprehensive psychological science as the substantive and methodological root of any educational or training enterprise in the field of psychology and without depreciating the value of scientist or scientist-professional training programs for certain specific objectives" (p. 442, original emphasis). The same phenomenon applies in other content-based training models, leading to the new understanding that relative emphasis is key in

the classifying content-based training models. Figure 1 illustrates the hypothetical relative position of content-based models along a continuum, with the practice component at one end and the science/research component at the other.

The number of repeating and overlapping items in the initial training-model scale is, to some degree, attributable to similar characteristics that different training models. For example, the of "Giving a wide range of practical experiences" is a feature of the SP model, but also typifies the practitioner (Korman, 1974) and the practitioner-scholar models (Bell & Hausman, 2014). Furthermore, "Teaching staff performing psychological practices" is required for the scientist-practitioner (Belar & Perry, 1992) and practitioner-scholar models (Bell & Hausman, 2014). "The use of scientifically based interventions" is one feature explicitly attributed to the scientist-practitioner (Belar & Perry, 1992) and CS models (McFall, 1991), but it implies use of a scientific approach in psychological practice. All training models in professional psychology advocate this, even if not explicitly stated. Learly, then, multiple training models share some characteristics, possibly leading to repetition of items representing these characteristics. This phenomenon provides evidence that classification of training models in professional psy-

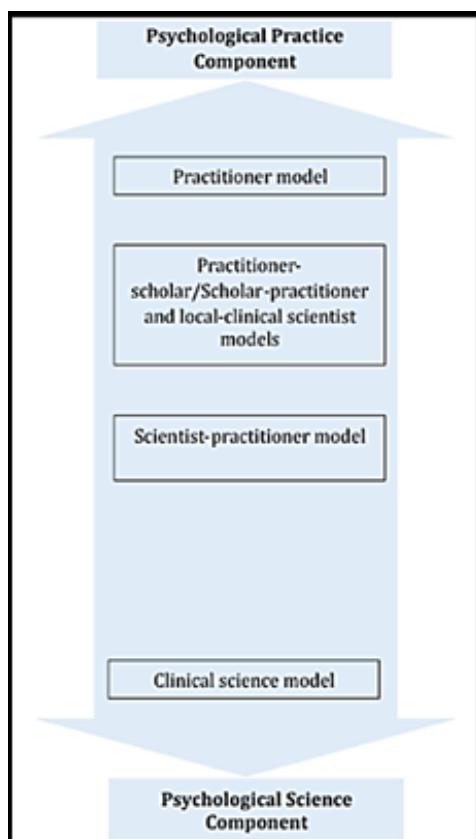


Figure 1. Content-based Models' Relative Position in the Range of Practice and Science Components in Professional Psychology Education

chology education is not mutually exclusive—a thought that previous researchers have expressed (Helmes, 2015, personal communication).

Results from this study, related to the validity and reliability coefficients, as well as qualitative comments provided by expert respondents have provided significant input applicable to improving the training-model scale. The training-model scale's final version consists of 77 items grouped into five clusters based on *the main components* of training-model classifications. This arrangement is different from the item grouping applied before the instrument testing, which was done on the basis of the *titles* of the training models (the Scientist-Practitioner/SP cluster, the Clinical-Science/CS cluster, and so on). The main training model's classification, the basis for item grouping in the new scale, include the following components: 1) practice; 2) science/research; 3) integration of science and practice; 4) local-clinical scientist; and, 5) competency. Items on the revised training-model scale were rated on a Likert scale with five alternative answers, starting from "Not at all" (1) to "Very high degree" (5), based on respondents' judgment of whether the stated condition applies in his/her professional program. The higher the score on a certain group of items, the higher the program's incidence of characteristics represented by the item cluster.

5. Conclusion

To conclude, further testing on the instrument should be undertaken to examine the revised training-model scale's psychometric properties for use in Indonesia. Further study should also provide an overview of the scale's applicability in contexts other than Indonesia. This is especially important when specific training-models' conception and/or application is not yet fully realized in professional psychology education.

Acknowledgements

This research was funded by the Indonesia Endowment Fund for Education (LPDP). The first author wishes to acknowledge LPDP for providing a scholarship to pursue a PhD degree in the Department of Psychology, College of Healthcare Sciences, James Cook University, Australia.

Declaration of Interest

The authors report no conflicts of interest in this work.

References

Aiken, L.R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955-959. doi: 10.1177/001316448004000419.

- Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131-142. doi: 10.1177/0013164485451012.
- Azwar, S. (2012). *Reliabilitas dan validitas/Reliability and validity* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Baker, D.B., & Benjamin, L.T., Jr. (2000). The affirmation of the scientist-practitioner: A look back at Boulder. *American Psychologist*, 55(2), 241-247. doi: 10.1037/0003-066X.55.2.241.
- BAN-PT. (2013a). *Buku 1: Naskah akademik*. Retrieved from <http://ppm.um-surabaya.ac.id/wp-content/uploads/2016/12/BUKU-1-Naskah-Akademik.pdf>.
- BAN-PT. (2013b). *Buku 2: Standar dan prosedur*. Retrieved from <http://ppm.um-surabaya.ac.id/wp-content/uploads/2016/12/BUKU-2-Standar-dan-Prosedur-Akreditasi.pdf>.
- Belar, C.D., & Perry, N.W. (1992). National conference on scientist-practitioner education and training for the professional practice of psychology. *American Psychologist*, 47(1), 71-75. doi: 10.1037/0003-066X.47.1.71.
- Bell, D.J., & Hausman, E.M. (2014). Training models in professional psychology doctoral programs. In W.B. Johnson & N.J. Kaslow (Eds.), *The Oxford handbook of education and training in professional psychology* (pp. 33-51). NY: Oxford University Press.
- British Psychological Society. (2015). *Standards for the accreditation of Doctoral programmes in clinical psychology*. Retrieved from http://www.bps.org.uk/system/files/Public%20files/PaCT/clinical_accreditation_2015_web.pdf.
- Canadian Psychological Association. (2001). *Mutual recognition agreement of the regulatory bodies for professional psychologists in Canada*. Retrieved from <http://www.cpa.ca/docs/File/MRA.pdf>.
- Canadian Psychological Association. (2011). *Accreditation Standards and Procedures for Doctoral Programmes and Internships in Professional Psychology*. Retrieved from http://www.cpa.ca/docs/File/Accreditation/Accreditation_2011.pdf.
- Cautin, R.L., & Baker, D.B. (2014). A history of education and training in professional psychology. In W. B. Johnson & N. J. Kaslow (Eds.), *The Oxford handbook of education and training in professional psychology* (pp. 17-32). NY: Oxford University Press.
- Cherry, D.K., Messenger, L.C., & Jacoby, A.M. (2000). An examination of training model outcomes in clinical psychology programs. *Professional Psychology: Research and Practice*, 31(5), 562-568. doi: 10.1037/0735-7028.31.5.562.
- Ellis, H.C. (1992). Graduate education in psychology: past, present, and future. *American Psychologist*, 47(4), 570-576. doi: 10.1037/0003-066X.47.4.570.
- Fouad, N.A., Grus, C.L., Hatcher, R.L., Kaslow, N.J., Hutchings, P.S., Madson, M.B., . . . Crossman, R.E. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology*, 3(4S), S5-S26. doi: 10.1037/a0015832.
- Haynes, S.N., Richard, D.C.S., & Kubany, E.S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247. doi: 10.1037/1040-3590.7.3.238.
- Horn, R.A., Troyer, J.A., Hall, E.J., Mellott, R.N., Coté, L.S., & Marquis, J.D. (2007). The scientist-practitioner model: A rose by any other name is still a rose. *American Behavioral Scientist*, 50(6), 808-819. doi: 10.1177/0002764206296459.
- Korman, M. (1974). National conference on levels and patterns of professional training in psychology: The major themes. *American Psychologist*, 29(6), 441-449. doi: 10.1037/h0036469.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. doi: 10.1111/j.1744-6570.1975.tb01393.x.
- Leong, F.T.L., & Zachar, P. (1991). Development and validation of the scientist-practitioner inventory for psychology. *Journal of Counseling Psychology*, 38(3), 331-341. doi: 10.1037/0022-0167.38.3.331.
- McEvoy, G.M., Hayton, J.C., Warnick, A.P., Mumford, T.V., Hanks, S.H., & Blahna, M.J. (2005). A competency-based model for developing human resource professionals. *Journal of Management Education*, 29(3), 383-402. doi: 10.1177/1052562904267538.
- McFall, R.M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist*, 44, 75-88.
- McFall, R.M. (2006). Doctoral training in clinical psychology. *Annual Review of Clinical Psychology*, 2(1), 21-49. doi: 10.1146/annurev.clinpsy.2.022305.095245.
- Merlo, L.J., Collins, A., & Bernstein, J. (2008). CUDCP-affiliated clinical psychology student views of their science training. *Training and Education in Professional Psychology*

Psychology, 2(1), 58-65. doi: 10.1037/1931-3918.2.1.58.

National Council of Schools and Programs of Professional Psychology. (2014). *Competency developmental achievement levels*. Retrieved from <http://www.thencspp.com/DALof%20NCSPP%209-21-07.pdf>.

Ningdyah, A.E.M., Greenwood, K.M., Kidd, G., Helmes, E.C., & Thompson, C. (2016). Training models in professional psychology education: A literature review. *Anima Indonesian Psychological Journal*, 31(4), 149-159.

Nixon, M. (1994). Practices and needs in psychological training: A survey of 28 countries. *Australian Psychologist*, 29(3), 166-173. doi: 10.1080/00050069408257345.

Pachana, N.A., O'Donovan, A., & Helmes, E. (2006). Australian clinical psychology training program directors survey. *Australian Psychologist*, 41(3), 168-178. doi: 10.1080/00050060600820644.

Rodolfa, E.R., Baker, J., DeMers, S., Hilson, A., Meck, D., Schaffer, J., . . . Webb, C. (2014). Professional psychology competency initiatives: Implications for training, regulation, and practice. *South African Journal of Psychology*, 44(2), 121-135. doi: 10.1177/0081246314522371.

Rodolfa, E.R., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice*, 36(4), 347-354. doi: 10.1037/0735-7028.36.4.347.

Rodolfa, E.R., Kaslow, N.J., Stewart, A.E., Keilin, W. G., & Baker, J. (2005). Internship training: Do models really matter? *Professional Psychology: Research and Practice*, 36(1), 25-31. doi: 10.1037/0735-7028.36.1.25.

Rosnow, R.L., & Rosenthal, R. (2002). *Beginning behavioral research: A conceptual primer (4th ed.)*. New Jersey, NJ: Pearson Education Inc.

Stricker, G. (1997). Are science and practice commensurable? *American Psychologist*, 52(4), 442-448. doi: 10.1037/0003-066X.52.4.442.

Stricker, G., & Trierweiler, S.J. (1995). The local clinical scientist: a bridge between science and practice. *American Psychologist*, 50(12), 995-1002. doi: 10.1037/0003-066X.50.12.995.

Sullivan, J., & Conoley, J.C. (2001). Best practices in the supervision of interns. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV*. Bethesda, MD: National Association of School Psychologists.

Universitas Surabaya. (2015). *Persyaratan pendaftaran magister psikologi profesi/admission requirements into master of professional psychology*. Retrieved from <http://psikologi.ubaya.ac.id/wp-content/uploads/2013/09/brosur-PPMPP-bagian-2-rev-yudi.jpg>.

Yang, W.-C. (2011). Applying content validity coefficient and homogeneity reliability coefficient to investigate the experiential marketing scale for leisure farms. *Journal of Global Business Management*, 7(1). Retrieved from <http://www.jgbm.org/page/8%20Wan-Chi%20Yang.pdf>.

Appendix

Table 1. Summary of Descriptive Statistics for Expert Evaluation on the Training Models Scale, Content Validity Coefficients (V), And Homogeneity Coefficients (H) (* p < 0.05, ** p < 0.01)

| SP Items | | | | | SP Items | | | | | SP Items | | | | |
|----------|------|------|--------|--------|----------|------|------|--------|--------|----------|------|------|--------|--------|
| No. | Mean | SD | V | H | No. | Mean | SD | V | H | No. | Mean | SD | V | H |
| 1 | 5.00 | 0 | 1.00** | 1.00** | 36 | 4.80 | 0.40 | 0.95** | 0.83* | 71 | 4.80 | 0.40 | 0.95** | 0.83* |
| 2 | 4.80 | 0.40 | 0.95** | 0.83* | 37 | 4.00 | 0.89 | 0.75 | 0.50 | 72 | 4.80 | 0.40 | 0.95** | 0.83* |
| 3 | 4.80 | 0.40 | 0.95** | 0.83* | 38 | 5.00 | 0 | 1.00** | 1.00** | 73 | 4.80 | 0.40 | 0.95** | 0.83* |
| 4 | 5.00 | 0 | 1.00** | 1.00** | 39 | 5.00 | 0 | 1.00** | 1.00** | 74 | 4.80 | 0.40 | 0.95** | 0.83* |
| 5 | 4.80 | 0.40 | 0.95** | 0.83* | 40 | 5.00 | 0 | 1.00** | 1.00** | 75 | 4.80 | 0.40 | 0.95** | 0.83* |
| 6 | 4.80 | 0.40 | 0.95** | 0.83* | 41 | 4.80 | 0.40 | 0.95** | 0.83* | 76 | 4.80 | 0.40 | 0.95** | 0.83* |
| 7 | 5.00 | 0 | 1.00** | 1.00** | 42 | 4.80 | 0.40 | 0.95** | 0.83* | 77 | 4.40 | 1.20 | 0.85* | 0.50 |
| 8 | 4.00 | 1.55 | 0.75 | 0.25 | 43 | 5.00 | 0 | 1.00** | 1.00** | 78 | 4.20 | 1.17 | 0.80* | 0.42 |
| 9 | 4.40 | 1.20 | 0.85* | 0.50 | 44 | 4.80 | 0.40 | 0.95** | 0.83* | 79 | 4.60 | 0.49 | 0.90* | 0.75* |
| 10 | 4.40 | 1.20 | 0.85* | 0.50 | 45 | 5.00 | 0 | 1.00** | 1.00** | 80 | 3.60 | 1.50 | 0.65 | 0.17 |
| 11 | 4.40 | 1.20 | 0.85* | 0.50 | 46 | 5.00 | 0 | 1.00** | 1.00** | 81 | 4.80 | 0.40 | 0.95** | 0.83* |
| 12 | 4.40 | 1.20 | 0.85* | 0.50 | 47 | 5.00 | 0 | 1.00** | 1.00** | 82 | 4.40 | 0.80 | 0.85* | 0.58 |
| 13 | 4.60 | 0.80 | 0.90* | 0.67 | 48 | 5.00 | 0 | 1.00** | 1.00** | 83 | 5.00 | 0 | 1.00** | 1.00** |
| 14 | 4.80 | 0.40 | 0.95** | 0.83* | 49 | 4.40 | 0.80 | 0.85* | 0.58 | 84 | 5.00 | 0 | 1.00** | 1.00** |
| 15 | 5.00 | 0 | 1.00** | 1.00** | 50 | 4.80 | 0.40 | 0.95** | 0.83* | 85 | 5.00 | 0 | 1.00** | 1.00** |
| 16 | 5.00 | 0 | 1.00** | 1.00** | 51 | 4.80 | 0.40 | 0.95** | 0.83* | 86 | 5.00 | 0 | 1.00** | 1.00** |
| 17 | 4.80 | 0.40 | 0.95** | 0.83* | 52 | 4.80 | 0.40 | 0.95** | 0.83* | 87 | 5.00 | 0 | 1.00** | 1.00** |
| 18 | 5.00 | 0 | 1.00** | 1.00** | 53 | 4.80 | 0.40 | 0.95** | 0.83* | 88 | 4.60 | 0.49 | 0.90* | 0.75* |
| 19 | 4.80 | 0.40 | 0.95** | 0.83* | 54 | 5.00 | 0 | 1.00** | 1.00** | 89 | 4.60 | 0.49 | 0.90* | 0.75* |
| 20 | 4.80 | 0.40 | 0.95** | 0.83* | 55 | 5.00 | 0 | 1.00** | 1.00** | 90 | 5.00 | 0 | 1.00** | 1.00** |
| 21 | 4.60 | 0.80 | 0.90* | 0.67 | 56 | 4.80 | 0.40 | 0.95** | 0.83* | 91 | 5.00 | 0 | 1.00** | 1.00** |
| 22 | 4.60 | 0.80 | 0.90* | 0.67 | 57 | 5.00 | 0 | 1.00** | 1.00** | 92 | 5.00 | 0 | 1.00** | 1.00** |
| 23 | 5.00 | 0 | 1.00** | 1.00** | 58 | 4.80 | 0.40 | 0.95** | 0.83* | 93 | 4.80 | 0.40 | 0.95** | 0.83* |
| 24 | 4.80 | 0.40 | 0.95** | 0.83* | 59 | 3.60 | 1.50 | 0.65 | 0.17 | 94 | 4.80 | 0.40 | 0.95** | 0.83* |
| 25 | 4.80 | 0.40 | 0.95** | 0.83* | 60 | 4.60 | 0.49 | 0.90* | 0.75* | 95 | 4.80 | 0.40 | 0.95** | 0.83* |
| 26 | 4.60 | 0.49 | 0.90* | 0.75* | 61 | 4.80 | 0.40 | 0.95** | 0.83* | 96 | 5.00 | 0 | 1.00** | 1.00** |
| 27 | 4.80 | 0.40 | 0.95** | 0.83* | 62 | 4.80 | 0.40 | 0.95** | 0.83* | 97 | 5.00 | 0 | 1.00** | 1.00** |
| 28 | 5.00 | 0 | 1.00** | 1.00** | 63 | 4.80 | 0.40 | 0.95** | 0.83* | 98 | 5.00 | 0 | 1.00** | 1.00** |
| 29 | 4.80 | 0.40 | 0.95** | 0.83* | 64 | 4.80 | 0.40 | 0.95** | 0.83* | 99 | 5.00 | 0 | 1.00** | 1.00** |
| 30 | 4.80 | 0.40 | 0.95** | 0.83* | 65 | 4.80 | 0.40 | 0.95** | 0.83* | 100 | 5.00 | 0 | 1.00** | 1.00** |
| 31 | 4.80 | 0.40 | 0.95** | 0.83* | 66 | 4.80 | 0.40 | 0.95** | 0.83* | 101 | 5.00 | 0 | 1.00** | 1.00** |
| 32 | 4.20 | 0.75 | 0.80* | 0.58 | 67 | 4.60 | 0.49 | 0.90* | 0.75* | 102 | 5.00 | 0 | 1.00** | 1.00** |
| 33 | 4.60 | 0.80 | 0.90* | 0.67 | 68 | 4.80 | 0.40 | 0.95** | 0.83* | 103 | 4.80 | 0.40 | 0.95** | 0.83* |
| 34 | 5.00 | 0 | 1.00** | 1.00** | 69 | 4.60 | 0.49 | 0.90* | 0.75* | 104 | 5.00 | 0 | 1.00** | 1.00** |
| 35 | 4.40 | 0.80 | 0.85* | 0.58 | 70 | 4.80 | 0.40 | 0.95** | 0.83* | 105 | 4.80 | 0.40 | 0.95** | 0.83* |

Table 1. Continued

| SP Items | | | | | SP Items | | | | | SP Items | | | | |
|----------|------|------|--------|--------|----------|------|------|--------|--------|----------|------|------|--------|--------|
| No. | Mean | SD | V | H | No. | Mean | SD | V | H | No. | Mean | SD | V | H |
| 106 | 4.40 | 0.49 | 0.85* | 0.75* | 120 | 4.80 | 0.40 | 0.95** | 0.83* | 134 | 5.00 | 0 | 1.00** | 1.00** |
| 107 | 4.60 | 0.49 | 0.90* | 0.75* | 121 | 4.80 | 0.40 | 0.95** | 0.83* | 135 | 4.80 | 0.40 | 0.95** | 0.83* |
| 108 | 4.40 | 0.49 | 0.85* | 0.75* | 122 | 4.80 | 0.40 | 0.95** | 0.83* | 136 | 4.40 | 0.80 | 0.85* | 0.58 |
| 109 | 4.80 | 0.40 | 0.95** | 0.83* | 123 | 4.80 | 0.40 | 0.95** | 0.83* | 137 | 4.80 | 0.40 | 0.95** | 0.83* |
| 110 | 5.00 | 0 | 1.00** | 1.00** | 124 | 4.80 | 0.40 | 0.95** | 0.83* | 138 | 5.00 | 0 | 1.00** | 1.00** |
| 111 | 5.00 | 0 | 1.00** | 1.00** | 125 | 4.60 | 0.49 | 0.90* | 0.75* | 139 | 5.00 | 0 | 1.00** | 1.00** |
| 112 | 4.80 | 0.40 | 0.95** | 0.83* | 126 | 4.60 | 0.49 | 0.90* | 0.75* | 140 | 4.40 | 0.80 | 0.85* | 0.58 |
| 113 | 5.00 | 0 | 1.00** | 1.00** | 127 | 4.40 | 0.80 | 0.85* | 0.58 | 141 | 4.80 | 0.40 | 0.95** | 0.83* |
| 114 | 5.00 | 0 | 1.00** | 1.00** | 128 | 4.80 | 0.40 | 0.95** | 0.83* | 142 | 5.00 | 0 | 1.00** | 1.00** |
| 115 | 4.80 | 0.40 | 0.95** | 0.83* | 129 | 5.00 | 0 | 1.00** | 1.00** | 143 | 4.80 | 0.40 | 0.95** | 0.83* |
| 116 | 5.00 | 0 | 1.00** | 1.00** | 130 | 5.00 | 0 | 1.00** | 1.00** | 144 | 4.80 | 0.40 | 0.95** | 0.83* |
| 117 | 4.80 | 0.40 | 0.95** | 0.83* | 131 | 4.60 | 0.49 | 0.90* | 0.75* | 145 | 4.80 | 0.40 | 0.95** | 0.83* |
| 118 | 4.80 | 0.40 | 0.95** | 0.83* | 132 | 5.00 | 0 | 1.00** | 1.00** | | | | | |
| 119 | 5.00 | 0 | 1.00** | 1.00** | 133 | 4.80 | 0.40 | 0.95** | 0.83* | | | | | |

| SP Items | | | | | LCS Items | | | | | LCS Items | | | | |
|----------|------|------|--------|--------|-----------|------|------|--------|--------|-----------|------|------|--------|--------|
| No. | Mean | SD | V | H | No. | Mean | SD | V | H | No. | Mean | SD | V | H |
| 146 | 5.00 | 0 | 1.00** | 1.00** | 151 | 4.80 | 0.40 | 0.95** | 0.83* | 157 | 5.00 | 0 | 1.00** | 1.00** |
| 147 | 5.00 | 0 | 1.00** | 1.00** | 152 | 5.00 | 0 | 1.00** | 1.00** | 158 | 4.80 | 0.40 | 0.95** | 0.83* |
| 148 | 4.80 | 0.40 | 0.95** | 0.83* | 153 | 5.00 | 0 | 1.00** | 1.00** | 159 | 4.80 | 0.40 | 0.95** | 0.83* |
| 149 | 4.80 | 0.40 | 0.95** | 0.83* | 154 | 5.00 | 0 | 1.00** | 1.00** | 160 | 4.60 | 0.49 | 0.90* | 0.75* |
| 150 | 4.80 | 0.40 | 0.95** | 0.83* | 155 | 4.80 | 0.40 | 0.95** | 0.83* | 161 | 4.80 | 0.40 | 0.95** | 0.83* |
| | | | | | 156 | 4.80 | 0.40 | 0.95** | 0.83* | 162 | 4.80 | 0.40 | 0.95** | 0.83* |

| CS Items | | | | | CS Items | | | | | CS Items | | | | |
|----------|------|------|--------|--------|----------|------|------|--------|--------|----------|------|------|--------|--------|
| No. | Mean | SD | V | H | No. | Mean | SD | V | H | No. | Mean | SD | V | H |
| 163 | 4.80 | 0.40 | 0.95** | 0.83* | 176 | 5.00 | 0 | 1.00** | 1.00** | 181 | 5.00 | 0 | 1.00** | 1.00** |
| 164 | 4.40 | 0.80 | 0.85* | 0.58 | 177 | 4.60 | 0.80 | 0.90* | 0.67 | 182 | 5.00 | 0 | 1.00** | 1.00** |
| 165 | 4.40 | 0.80 | 0.85* | 0.58 | 178 | 5.00 | 0 | 1.00** | 1.00** | 183 | 4.00 | 2.00 | 0.75 | 0.17 |
| 166 | 4.60 | 0.80 | 0.90* | 0.67 | 179 | 4.80 | 0.40 | 0.95** | 0.83* | 184 | 5.00 | 0 | 1.00** | 1.00** |
| 167 | 4.80 | 0.40 | 0.95** | 0.83* | 180 | 4.80 | 0.40 | 0.95** | 0.83* | 185 | 5.00 | 0 | 1.00** | 1.00** |
| 168 | 4.40 | 0.80 | 0.85* | 0.58 | | | | | | 186 | 4.60 | 0.49 | 0.90* | 0.75* |
| 169 | 4.80 | 0.40 | 0.95** | 0.83* | | | | | | 187 | 4.80 | 0.40 | 0.95** | 0.83* |
| 170 | 4.60 | 0.49 | 0.90* | 0.75* | | | | | | 188 | 4.80 | 0.40 | 0.95** | 0.83* |
| 171 | 4.60 | 0.80 | 0.90* | 0.67 | | | | | | 189 | 4.60 | 0.49 | 0.90* | 0.75* |
| 172 | 5.00 | 0 | 1.00** | 1.00** | | | | | | 190 | 4.80 | 0.40 | 0.95** | 0.83* |
| 173 | 5.00 | 0 | 1.00** | 1.00** | | | | | | 191 | 5.00 | 0 | 1.00** | 1.00** |
| 174 | 4.60 | 0.49 | 0.90* | 0.75* | | | | | | 192 | 5.00 | 0 | 1.00** | 1.00** |
| 175 | 4.60 | 0.49 | 0.90* | 0.75* | | | | | | 193 | 4.80 | 0.40 | 0.95** | 0.83* |
| | | | | | | | | | | 194 | 4.80 | 0.40 | 0.95** | 0.83* |
| | | | | | | | | | | 195 | 5.00 | 0 | 1.00** | 1.00** |

Table 2. Comments of Expert Respondents to Items in Particular and the Scale in General

| Types of comments | | Comments |
|-------------------|--------|--|
| Specific related) | (Item- | <ul style="list-style-type: none"> - It is not necessary to include some subjects provided in undergraduate level courses in the professional program (Part 2A, item no. 1-53; R1). - Check the words highlighted and change some of the words currently in use into suggested alternative words in Indonesian, as follows (Part 2A; R2): - No. 19: "tervalidasi" to be changed to "berbasis ilmiah" - No. 24: "penyakit" to "gangguan" - No 26: "kehidupan praktek" to "praktek psikologi" - No 32: "pelayanan" to "pelayanan psikologis" - No 37: "praktikum penelitian" to "praktek penelitian" - No 59" "praktek dasar" to "dasar-dasar praktek" - No 60: "aktivitas praktek" to "pendampingan praktek/magang" - No 82: "beragam" to "berbagai metode" - No 120: "lokal" to "sesuai kultur budaya lokal" - No 131: "tujuan" to "visi dan misi" - No 136: "kurang beruntung" to "marginal" - No 160: "kekuasaan, otoritas" to "kaum berkuasa" |
| General | | <ul style="list-style-type: none"> - Too many items; items are overlapping. Some items that have appeared before are subsequently repeated using different wording (R1). - Too many items, and sentences used are too long (R3). |

Note: R= respondent; example: R1= respondent 1