



Validation of the Supervision Outcome Scale in Counselor Professional Education in Indonesia: A Rasch Model Approach

Afriyadi Sofyan^{1,2} , Nur Hidayah¹ , Bambang Budi Wiyono¹ , M. Ramli¹ , Yuying Tsong

¹Universitas Negeri Malang, Malang, Indonesia

²Universitas Negeri Semarang, Semarang, Indonesia

³California State University, California, United States of America

afriyadi.sofyan.2201119@students.um.ac.id

Abstract. This study investigates the applicability of the Supervision Outcome Scale (SOS), initially developed by Yuying Tsong, in the context of counselor professional education (CPE) in Indonesia. The research objective was to validate the Indonesian SOS (ID-SOS) version using the Rasch Rating Scale Model. The participants comprised 133 students from CPE programs across five universities in four Indonesian provinces. They were evaluated on their acceptance of the supervision outcomes through the ID-SOS, which included seven items tailored to measure clinical and multicultural competencies. The Rasch model analysis assessed item fit, internal consistency, and reliability, revealing a person reliability of 0.80 and an item reliability of 0.94. The findings suggest a robust psychometric structure of the ID-SOS, demonstrated by the high percentage of variance explained by the measures (77.1%). This validation confirms the ID-SOS's potential as a robust tool for measuring supervision efficacy in Indonesian counselor education, offering significant implications for enhancing the quality and effectiveness of counseling supervision in diverse educational settings.

Keywords: Supervision Outcome Scale (SOS), Counselor Professional Education (CPE), Rasch Rating Scale Model, Clinical and Multicultural Competence, Counselor Education in Indonesia.

1 Introduction

In the counselor professional education (CPE) program, there are two activities that Hodges [1] calls the backbone of counseling education programs in general. The two activities referred to are practicum and internship activities, which, in implementation by the Council for Accreditation of Counseling and Related Educational Programs (CACREP), are required to be carried out by prospective counselors for at least 100 hours for practicum activities and 600 hours for internship activities under a measurable supervision process by a competent or professional supervisor [2].

© The Author(s) 2024

Q. Zhang (ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 23, https://doi.org/10.2991/978-94-6463-494-5_14

Supervision in a counseling context is a collaborative and evaluative process in which a more experienced practitioner (supervisor) provides guidance, feedback, and support to a more junior practitioner or student (supervisee) [3], to increase professional competence, ensuring the delivery of ethical and effective counseling services [4, 5], as well as supporting the personal and professional growth of supervision [6]. In counseling internship activities, supervision plays a vital role as an integral part of the learning process and development of students' practical skills in actual professional settings [7].

Measurable supervision in the supervision process contributes to improving quality, understanding, and professional growth by implementing effective supervision practices, especially in counseling internship work practices [8-10]. The supervision carried out in practicums and internships in counselor professional education is specific supervision, namely clinical supervision. Supervision generally covers various aspects of work, such as management, communication skills, and general performance improvement. Moreover, clinical supervision is focused on developing counseling skills, understanding cases, selecting appropriate interventions, and applying ethics and professional counseling standards [11-13].

In practicums and internships, professional education students will be accompanied by at least a field supervisor and a faculty or university supervisor who has been certified as a supervisor and has practical experience as a counselor for at least four years [1, 7]. Clinical supervisors are expected to be able to act as mentors, supervisors, and mentors for prospective counselors who are in the process of improving their competence [11]. Supervisors are expected to help prospective counselors develop skills, ensure the quality of counseling services provided, and ensure that counselors work within established ethical boundaries and professional standards.

Although professional counselor education programs in Indonesia have been running for more than 20 years, in 1999/2000, initial pilots began at Padang State University [14], and in 2007, at Semarang State University. Unfortunately, until now, a valid scale for measuring the results of counseling internship supervision for professional counselor education students has not yet been implemented. Research and publications on supervision and the scale used still need to be improved. Meanwhile, researchers who have attempted to study supervision in the counseling field in Indonesia, for example, Hafina [15], whose research focuses on developing technology (web-based) programs to evaluate and supervise counseling skills, are limited to practicum activities and carried out by undergraduate guidance and counseling study program students. Furthermore, Taufiq [16] developed the supervisory work alliance scale by testing it using the Rasch model. This instrument could still not measure the supervisory work alliance variable because the results exceeded the measurement standards of unidimensionality. The scale he tested was limited to 75 school counselors undertaking an internship in the school counselor in-service training education program.

The supervision outcome scale (SOS) is essential and a significant need in counselor professional education to provide a structured and objective evaluation of the efficacy and outcomes of the supervision process. The SOS provides empirical data to improve

the quality of supervision and counseling practice. Yuying Tsong [17] developed the SOS and reported its psychometric properties in 2014. She conducted research with exploratory and confirmatory factor analysis on two groups of doctoral students in the fields of counseling and clinical psychology, which showed that the SOS succeeded in identifying two different dimensions related to the impact of supervision, namely the impact on clinical competence (including reducing symptoms in clients and increasing supervisor competence) and the impact on multicultural competence (increasing multicultural abilities in supervisors). Furthermore, SOS developed by Tsong was proven to have internal reliability and good concurrent validity, characterized by a significant correlation with working alliances in supervision

2 Research Questions and Objectives

Our research sought to confirm whether the SOS developed by Yuying Tsong [17] is appropriate in the Indonesian context during counseling internships for counselor professional education students in Indonesia. This study aims to validate the Indonesian SOS (ID-SOS) version using the Rasch model.

3 Methodology

The research validated the ID-SOS with analysis carried out using the Rasch Rating Scale Model [18]. Proposed measurement properties to evaluate ID-SOS via Rasch analysis: Misfit Item; Internal consistency reliability and separation indices; unidimensionality and Local dependency; precision measurement, & rating scale diagnostic-data analysis using the Rasch model (WINSTEPS 5.1.7.0).

This study involved 133 respondents aged 25 - 58 who were CPE program students from five universities in four provinces in Indonesia (North Sumatra, West Sumatra, Central Java, and Bali). The CPE students were recruited from counselors working in junior high and senior high schools with at least two or more years of counseling experience.

Table 1. Demographic data of counselor trainee (N = 133).

Demography		Frequency	Percentage (%)
University	University 1	19	14
	University 2	48	36
	University 3	38	28
	University 4	19	14
	University 5	9	7
Years of service	≥10 years	22	16
	7-9 years	34	26
	4-6 years	53	40
	1-3 years	24	18

4 Result

From Table 2, we can understand that the reliability value is 0.80 for the person, which indicates that the person's measurement is relatively consistent. For items, the reliability value is higher, namely 0.94, which shows that these items provide consistent results in measuring the desired construct. Separation Index: The separation index measures how well a set of items differentiates between respondents with different ability levels (for people) or how well a set of respondents differentiates between items with different difficulty levels (for items). A value of 1.99 for the person indicates some separation, while a value of 3.81 for items indicates a more precise separation between items based on difficulty.

For the person, the mean measure is 7.64, while for items, it is 0.00. In Rasch models, the mean value for items is usually set to zero so that the mean value for persons can be interpreted as well as being above the average difficulty of the items.

The high alpha value (0.94) for the items indicates that the items as a whole have high homogeneity in measuring the same construct. The percentage of raw variance explained by measures calculated through principal component analysis (PCA) shows a value of 77.1% for the item, indicating that most of the variability in the data can be explained by the Rasch model, which indicates that the model is quite good at capturing the information contained by the data.

Table 2. Summary statistics of person and item (I =7, N = 133).

	Reliability	Separation index	Mean measure*)	Cronbach's alpha	Raw variance explained by measures**)
Person	0.80	1.99	7.64		
Item	0.94	3.81	0.00	0.94	77.1%

*) Measure in logit

**) Computed via PCA.

Table 3 shows that for the ID-SOS answer option, which reads "Not help at all," not a single respondent chose it from all the statements in ID-SOS. Furthermore, the logit value for each answer choice for the other ID-SOS statement differs from a small logit value for the answer choice with the minimum score to the logit value for the large answer choice with the maximum score. This finding indicates the respondents' ability to verify the differences between various ID-SOS answer options. The observation level associated with the Andrich Threshold also changes monotonically from NONE, then moves towards negative logit and increases towards positive logit (12.39 logit)

for each answer choice. These results also confirm that the answer choices are valid and do not mislead respondents.

Table 3. Item threshold and fit indices of response format (I =7, N = 133).

Category Label	Andrich Threshold	Observed Average	Observed Count (%)	Infit	Outfit
Not help at all	-	-	-	-	-
Helpful, but very little	NONE	-13.96	2	0.66	0.16
Somewhat helpful	-12.62	-5.29	10	1.19	0.26
Very helpful	0.23	7.91	63	0.99	0.82
Extremely helpful	12.39	13.56	25	0	0.51

Table 4 presents statistical measures for ID-SOS, which explain in detail the size of the items, the level of suitability of the items to produce measurement productivity (infit and outfit MNSQ), the level of accuracy of the resulting measurements (S.E. Model), and item discrimination (Point Measure Correlation). As shown in Table 4, Item 5¹ and 6² is the most difficult item (1.35 logit) for all respondents. Meanwhile, Item 3³ reaches the lowest level of difficulty (-2.07 logit).

In this study, almost all ID-SOS items have an Outfit MNSQ between $0.5 < \text{MNSQ} < 1.5$; only item 3 has an Outfit MNSQ below 0.5 (0.2); however, Item 3 has an Outfit ZSTD -2.1, with a total of N = 133, which refers to the standard Z-fit value (ZSTD) being still acceptable in the range $-2.0 < \text{ZSTD} < +2.0$ [19], which indicates that the data fits the model. The "S.E Model" column provides each item's "Standard Error of Measure" information. The standard error in the proposed instrument is in the range of < 0.5 , which indicates decisive results produced by ID-SOS. In detail, as shown in Table 4, the Item Standard Error estimates ranged between 0.28 and 0.32. The data demonstrate reliable item fit and solid item capacity to express measurement precision. Based on the value shown in the Pt. Measure column, no negative values were identified in the Measure Corr column.

¹ Item 5: in english (Your multicultural counseling skills (e.g., skills that areculturally appropriate in working with diverse clients), in Indonesian (keterampilan konseling lintas budaya anda (misalnya, keterampilan yang sesuai secara budaya dalam bekerja dengan konseli yang beragam))

² Item 6: in English (Your multicultural beliefs/attitudes/awareness (e.g., awareness of your own worldviews)), in Indonesian (keyakinan/sikap/kesadaran lintas budaya Anda (misalnya, kesadaran akan pandangan tentang dunia anda sendiri)).

³ Item 3: in english (Your Counseling skill), in Indonesian (keterampilan konseling anda)

Table 4. The summary of item measure (I =7, N = 133).

Item	Total Score	Measure	S.E Model	infit		outfit		Pt. Measure
				MNSQ	ZSTD	MNSQ	ZSTD	
5	530	1.35	0.28	1.1	0.5	0.5	-0.8	0.8
6	530	1.35	0.28	0.8	-1.4	0.4	-1.2	0.8
7	538	0.74	0.27	0.8	-2.1	0.4	-0.9	0.9
2	550	-0.13	0.27	1.2	1.3	0.7	-0.1	0.8
1	551	-0.2	0.27	1.5	3.3	1.1	0.4	0.8
4	562	-1.05	0.29	0.8	-1.1	0.5	-0.7	0.9
3	573	-2.07	0.32	0.6	-2.2	0.2	-2.1	0.9

In addition to item measures, we present the person measures to assess supervision outcomes. The measurement results are shown in Table 5. Based on the Rasch calculations obtained, Table 5 presents the top five and bottom five responses from the 133 respondents who participated in this research. The highest person size (14.6 logit; S.E = 1.14) represents the highest acceptance of outcome supervision, obtained by respondent Number 6, 52, 59, 89, and 120. On the other hand, respondent Number 131 had the lowest acceptance of outcome supervision (-14.87 logit; S.E = 1.18) compared to all respondents involved in this research.

Based on item and person parameters estimation results, we apply Wright maps to plot the distribution of respondents' item abilities and difficulties on the "Rasch Ruler" with the same logit. The Wright-map for ID-SOS is presented in Figure 1. Figure 1 also illustrates the ID-SOS procedure in assessing supervision outcomes. Through this mechanism, we intend to compare the distribution pattern of respondents on the Wright map with the distribution of items based on Rasch calculations. Comparisons need to be made. However, people and items have the same units, and the logit is a linear interval unit [20].

Table 5. The summary of person measure (I =7, N = 133).

Person Entry Number	Total Score	Measure	S.E Model	infit		outfit		Pt. Measure
				MNSQ	ZSTD	MNSQ	ZSTD	
6	34	14.6	1.14	1.3716	0.7214	1.7655	0.9118	-0.0693
52	34	14.6	1.14	0.8164	-0.0592	0.4646	-0.0095	0.4699
59	34	14.6	1.14	0.8164	-0.0592	0.4646	-0.0095	0.4699
89	34	14.6	1.14	0.8164	-0.0592	0.4646	-0.0095	0.4699
120	34	14.6	1.14	1.3569	0.7014	1.6518	0.8617	-0.0437
220	21	-6.27	4.89	0.0057	-1.27	0.0039	-1.32	-0.0001
187	19	-11.41	0.92	0.8256	-0.3392	0.6089	-0.2394	0.5643
74	18	-12.2	0.87	0.476	-1.6995	0.4201	-1.1496	0.8474
99	17	-12.95	0.88	0.5895	-1.0894	0.4969	-1.0495	0.7984
131	15	-14.87	1.18	1.4771	0.8315	1.3096	0.6713	0.0693

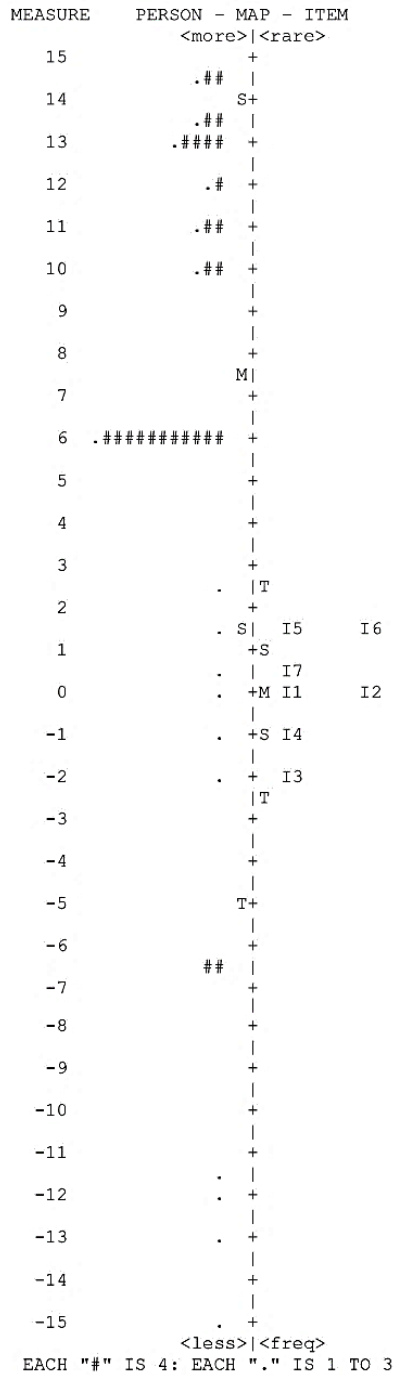


Fig. 1. Wright map

5 Conclusion

There remains a significant need for valid instruments to measure supervision outcomes in Indonesia, an area with limited research. Yuying Tsong developed the SOS, highlighting the importance of the scale in measuring the impact of supervision on clinical and multicultural competence. This development substantially enhances the quality of supervision and counseling practice through robust empirical data. Efforts to validate the SOS in its Indonesian version during the counseling internship activities for counselor education students in Indonesia have produced a reliable alternative. The findings indicate that the ID-SOS, comprising seven items, demonstrates solid and consistent psychometric properties, confirming its validity and reliability as a research instrument. Consequently, academics, researchers, and implementers of counselor professional education programs can use the ID-SOS in scientific studies related to the supervision of counselor education, particularly in Indonesia.

Acknowledgement. We want to acknowledge and thank (1) Center of Higher Education Fund (Balai Pembiayaan Pendidikan Tinggi/BPPT), The Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, and (2) Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP), Ministry of Finance Indonesia for granting a scholarship to the first author and the presenter to complete his study (Registration Number: B3172022010731235347 and BPI Number: 202209090667).

Acknowledgments. A third level heading in 9-point font size at the end of the paper is used for general acknowledgments, for example: This study was funded by X (grant number Y).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

- 1 Hodges, S.: 'The counseling practicum and internship manual: A resource for graduate counseling students' (Springer Publishing Company, 2019. 2019)
- 2 Borders, L.D.: 'Snapshot of clinical supervision in counseling and counselor education: A five-year review', *The Clinical Supervisor*, 2006, 24, (1-2), pp. 69-113
- 3 Bernard, J.M., and Goodyear, R.K.: 'Fundamentals of clinical supervision' (Allyn & Bacon, 2004. 2004)
- 4 Garrett, M.R.: 'Clinical Counseling Supervisors' Perspectives of Gatekeeping During Prelicensure Supervision', Walden University, 2023
- 5 Aydin, F., and Odaci, H.: 'Novice Counsellor Supervisees' Satisfaction with the ICP Supervision: Adapting the Supervisory Satisfaction Questionnaire Supervisee Form to Turkish', *International Journal for the Advancement of Counselling*, 2023, pp. 1-23
- 6 Terjesen, M., Ford, A.J., Allen, K., Lin, K., Yellico, M., and Walsh, O.A.: 'Supervision in Cognitive Behavioral Therapy: Theory, Practice, Culture, and Ethics': 'Handbook of Training and Supervision in Cognitive Behavioral Therapy' (Springer, 2023), pp. 3-27

- 7 Baird, B.N., and Mollen, D.: 'The internship, practicum, and field placement handbook: A guide for the helping professions' (Taylor & Francis, 2023. 2023)
- 8 Kalkan, B., and Nesime, C.: 'Supervision in counselor education: Exploration of current status and standards in Turkey', *Adiyaman University Journal of Educational Sciences*, 2019, 9, (2), pp. 271-290
- 9 Neyland-Brown, L., Laux, J.M., Reynolds, J.L., Kozlowski, K., and Piazza, N.J.: 'An Exploration of Supervision Training Opportunities for School Counselors', *Journal of School Counseling*, 2019, 17, (1), pp. n1
- 10 Wambu, G.W., and Myers, C.E.: 'School Counselor Site Supervisors' Perceptions of Preparedness and Training Needs', *Journal of School Counseling*, 2019, 17, (14), pp. n14
- 11 Corey, G., Haynes, R.H., Moulton, P., and Muratori, M.: 'Clinical supervision in the helping professions: A practical guide' (John Wiley & Sons, 2020. 2020)
- 12 Bernard, J.M., & Goodyear, R. K.: 'Fundamentals of Clinical Supervision Sixth Edition' (Pearson, 2019. 2019)
- 13 Garman, N.: 'The dream of clinical supervision, critical perspectives on the state of supervision, and our long-lived accountability nightmare', *Journal of Educational Supervision*, 2020, 3, (3), pp. 7
- 14 Habsy, B.A.: 'Filosofi ilmu bimbingan dan konseling Indonesia', *JP (Jurnal Pendidikan): Teori dan Praktik*, 2017, 2, (1), pp. 1-11
- 15 Hafina, A.R., Nandang Ramdhani, Rina Nurhudi Guntara, Rangga Gelar: 'Application Of Evaluation and Supervision Model Of Counselling Skills: Implementation, Testing, and Integration', 2022
- 16 Taufiq, A., Yudha, E.S., Md, Y.H., and Suryana, D.: 'Examining the Supervision Work Alliance Scale: A Rasch Model Approach', *The Open Psychology Journal*, 2021, 14, (1)
- 17 Tsong, Y., and Goodyear, R.K.: 'Assessing supervision's clinical and multicultural impacts: The Supervision Outcome Scale's psychometric properties', *Training and Education in Professional Psychology*, 2014, 8, (3), pp. 189
- 18 Andrich, D.: 'Rasch rating-scale model': 'Handbook of item response theory' (Chapman and Hall/CRC, 2016), pp. 75-94
- 19 Sumintono, B., and Widhiarso, W.: 'Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)' (Trim Komunikata Publishing House, 2014. 2014)
- 20 Boone, W.J., Townsend, J.S., and Staver, J.R.: 'Utilizing multifaceted Rasch measurement through FACETS to evaluate science education data sets composed of judges, respondents, and rating scale items: An exemplar utilizing the elementary science teaching analysis matrix instrument', *Science Education*, 2016, 100, (2), pp. 221-238

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

