INTER RATER RELIABILITY AS QUALITY ASSURANCE
OF KFZA TRANSLATION INTO INDONESIAN LANGUAGE

Ratri Atmoko Benedictus
Human Factor Engineering Dept. Hochschule Niederrhein, Germany
Faculty of Psychology, Atma Jaya Catholic University of Indonesia
ratrie.benedictus@atmajaya.ac.id

Abstract

Translation of a psychological measuring instrument is one of the harmonization efforts towards a cultural context, especially in terms of language. The process is not just utilizing the sophistication of translator technology, but also need to do an evaluation of the accuracy of a translation. This article aims to measure the degree of quality of early translation accuracy of a psychological measuring instrument, i.e. KFZA that measures work stress. KFZA itself stands for Kurz-Fragebogen zur Arbeitsanalyse that is developed by Prumper, Hartmannsgruber and Frese in the German language. The result indicated that the coefficient of Cohen's kappa (k) is .458 and it represents a weak of agreement between the two raters, although it is reliable at the 0.05 level. Since initial translation is a critical step of psychological test translation and adaptation, competence qualifications of each rater and the equivalence of competencies between raters are very important.

Keywords: Inter rater Reliability, Psychological Test, Translation, KFZA,

Introduction

The term "cultural blind" initially delivered by Triandis (2009) to describe the assumption that until 1960s all psychological discoveries are independent of culture. Further, Triandis then illustrated the concept of Cross-Cultural Psychology that emerged in early 1970s as the encounter of the collectivists against individualists. Prior to the 1960s, the development of psychology in collective countries, such as Indonesia, were strongly influenced by the results of psychological research conducted by developed countries with an individualist cultural background. Dialogue about cross-cultural psychology then began to grow rapidly along with the development of its research and recently with the concept of indigenous psychology. Fifty years later, Survey of Gabreya (2009) on one cross cultural psychology forum showed that the highest degree of satisfaction is on the networking activity (86%) and one of the lowest is at the follow-up action (46%), particularly in the development of scientific methods in psychology in the context of collectivist country.

Hofstede (n.d) defined Indonesia, with a low score of individualism, as a collectivist society and those dimension is the lowest score of the other six cultural dimensions. It means that people of Indonesia are fluent in use the term of "we" than "I". Moreover, they also has a high preference to engage with a certain social framework in which conformity to the ideals of society and the in-groups to which they belong is a very important thing. But this uniqueness does not seem to be a concern, even in the application of various kinds of psychological measurements in Indonesia. Globalization and rapid development of the technology has further worsen that condition. Psychological measurements that developed in the context of individualist countries are only translated into the aspect of its language and then applied to Indonesian society who is collectivist.

Chapman and Carter (1979) strongly argued that proper procedures need to be executed. If not, such studies, particularly Psychology, would suffer an error on its result due to more in its translation rather than differences in people or variables

being measured. Chapman and Carter also explained three resources which related to the problems on the across cultural measurement. First, short history of empirical research in developing countries resulted less number of measurement instrument which specifically developed for local setting. Second, many scholars from developing countries gained their high education graduation abroad who even use their familiar instrument for their home country with less consideration with the local setting. And third, universities, mostly in developing countries, trained their students in evaluation and research and even considered that a research being sophisticated since based on the previous foreign research. An Indonesian scholar, Suwartono (2016), criticized on the purchasing and procurement of the testing instruments from abroad that being simply translated into Indonesian. She further recommended that the process of translating a test tool not only focus on the issue of language, but also non-language, such as culture and the relevance of knowledge about the target population of the test.

International Test Commission actually had issued guidance for the process of across cultural test adaptation (International Test Commission, 2005). Seven guidelines have been developed to ensure the quality of the adapted test, so that it does not reduce the purpose of the measurement and the value of its benefits. That seven guidelines are guidelines on test use, guidelines on computer-based testing, guidelines on quality control, guidelines on test security, guidelines on test disposal, test-taker's guide, and statement on using tests for research.

This article attempted to provide how to measure the degree of quality in translating process of KFZA, as a tool for work stress measurement, from German language into Indonesian context. Work stress itself is not a new theme in research and study of occupational psychology in Indonesia. The database of psychological journal under High Education Directorate of Republic of Indonesia showed that scholars mostly refer to western theory following its measurements about work stress, such as Moorhead and Griffin (Bachroni & Asnawi, 1999), Behr and Newman (Sagala, 2013; Mahardayani & Dhania, 2014; Wulandari & Wibowo, 2013), Selye (Hidayati, Purwanto, & Yuwono, 2008), Palmer, Cooper and Thomas (Mahdia, 2014), Robbins (Safaria, 2011), NIOSH (Widhiastuti, 2002), Spielberger (Widodo & Djalali, 2011), Cooper and Straw (Prasetyo & Nurtjahjanti, 2012; Prasetyo, Nurtjahjanti, Fauziah & Kustanti, 2016), and Greenberg (Lubis & Budiman, 2014). There is only one group of researchers, ie. Apreviadezy & Puspitacandri (2014), who use Indonesian literature to design work stress measurements in their research. They use the Hardjana approach that measures work stress based on four components, namely: physical, intellectual, emotional, and behavioral. Hardjana's description about work stress was published by Kanisius publishers in 1994 in a popular Indonesian book entitled *Stress tanpa Distress* (Stress without distress). This became interesting because those title was initially introduced by Hans Selye in 1974 which published by Philadelphia publisher in New York.

Theoretical Review
Inter rater reliability as quality measurement in qualitative data analysis

In determining the quality of a measurement or how well a test captures the essence of a particular aspect, Howitt and Cramer (2011) pointed three criteria need to be considered, namely objectivity, reliability and validity. Objective mean the test or measure should provide similar outcomes irrespective of who is administering the test, while reliable might be defined as a consistency of the test at different points in time or across different circumstance. And last, valid refers to the extent to which a test assesses what it is claimed to measure. In content analysis

Tinsley and Weiss (1975) introduced the terminology of rater as person who make a judgment about some characteristic of an object by assigning it to some point on a scale defined in terms of that characteristic. In the qualitative research, Patton (2015) highlighted the argument of Janet Morse that inter rater reliability may be acceptable when everyone is asked the same question in the same way, but in the more adaptive, personalized, and flexible approach of interview guide and conversational interviewing.

Using percentage of agreement is the classic way to measure inter rater reliability, although it is good only for a view number of categories and rater in which some agreement is sure to occur (Iacobucci, 2001). Otherwise, the more random agreement is likely to occur, thus measuring level of reliability is better. Tinsley and Weiss (1975) explained the difference between agreement versus reliability. Inter rater agreement represents the different rater who tend to make the same judgment or exactly the same values about rated subject. Otherwise, inter rater reliability represents the degree to which the ratings of different raters are proportional when emerged as deviations from their means. Furthermore, inter rater reliability usually utilizes the indexes of correlation or analysis of variance. Table 1 shows the different levels of inter rater agreement versus inter rater reliability.

Table 1. Illustrating Different Levels of Inter rater Agreement and Inter rater Reliability for Interval-Scaled Data

Counselors	Case 1: High inter rater agreement and high inter rater reliability			Case 2: Low inter rater agreement and high inter rater reliability			Case 3: High inter rater agreement and low inter rater reliability		
	Rater			Rater			Rater		
	1	2	3	1	2	3	1	2	3
A	1	1	1	1	3	5	5	4	4
B	2	2	2	1	3	5	5	4	3
C	3	3	3	2	4	6	5	4	5
D	3	3	3	2	4	6	4	4	5
E	4	4	4	3	5	7	5	4	3
F	5	5	5	3	5	7	5	5	4
G	6	6	6	4	6	8	4	4	5
H	7	7	7	4	6	8	5	5	4
I	8	8	8	5	7	9	4	5	3
J	9	9	9	5	7	9	5	5	5
Mean	4.8	4.8	4.8	3.0	5.0	7.0	4.7	4.4	4.1
SD	2.7	2.7	2.7	1.5	1.5	1.5	.5	.5	.9

(Resource: Tinsley and Weiss, 1975)

Translation Process

Wild et al. (2005) presented ten stages in processing a translation and cultural adaptation which resulted from a TCA (Translation and Cultural Adaptation) working group within ISPOR's Quality of Life Special Interest Group on develop the Patient-Reported Outcome measures. Table 2 shows its activities for each step for translating and performing cultural adaptation.

Table 2. Ten Stages on translation and cultural adaptation process

No.	Stage	Component of Activities
1.	Preparation	Obtain permission and authorized to use the copyright material, develop explanation of concepts in the instrument to eliminate the misinterpretation of its items or concepts.
2.	Forward Translation	Develop at least two independent translations to eliminate the dependency to one person's own style of writing, provide the explanation of the concepts mainly to the forward translators.
3.	Reconciliation	Develop a consensus among the forward translations and resulted into single forward translation.
4.	Back Translation	Translate the single forward translation into source language to ensure the quality of translation.
5.	Back Translation Review	Review of the back translations against the source language to ensure its two equivalences.
6.	Harmonization	Provide harmonization between the new translation and the source version in order to detect and deal with its translation discrepancies.
7.	Cognitive Debriefing	Asses the level of comprehensibility and cognitive equivalence of the new translation.
8.	Review of result	Review the result of cognitive debriefing and finalized the translation, even in choosing of words or phrase.
9.	Proofreading	Proofread the final translation to eliminate the error of spelling, grammatical and others.
10.	Final Report	Write a report on the development of translation to obtain the clear reason about wording choices throughout the translation process.

(Resource: Wild et al., 2005)

Beaton et al. (2007) then simplified into six stages as a recommendation on cross-cultural adaptation. The first stage is initial translation which better to provide at least two forward translation from the original language to the target language. Those independent forward translation are produced by two bilingual translator who have the target language as their mother tongue. The second stage is to synthesize the forward translation with adding a third unbiased person to the team. He has a responsibility as a mediator within discussions of translation differences. Stage three is back translation which translate the newest version of translation into the original language by minimum two translator who totally blind to the original version. Stage four is expert committee that involve at least one each of a methodologist, professional on the topic which being measured, language professional, as well as all translators. The decision should be obtained on this stage to achieve the equivalence in four areas, namely semantic equivalence, idiomatic equivalence, experiential equivalence, and conceptual equivalence. Stage five is test of the pre-final version by using a field test of the new questionnaire. Ideally the field test uses between 30 and 40 persons from a target setting. And the last stage is documentation writing.

Concept of KFZA (Kurz-Fragebogen zur Arbeitsanalyse)

On this paper, instrument for work stress measurement refers to the concept KFZA or *Kurz-Fragebogen zur Arbeitsanalyse*. Prumper, Hartmannsgruber and Frese (1995) generated KFZA from the compilation of several tools that measure

the stress dimensions of the job, such as Instrument for Stress-Related Job Analysis by Semmer, Instrument for Stress-Related Job Analysis of Computer Work by Zapf, Questionnaire on Stress-Conditions at Work by Frese, Job Diagnostic Survey by Hackman and Oldham, Subjective Work Analyzing by Udris and Alioth, and Working Climate by Reosensitel et al. Table 3 presents the test construction of KFZA.

Table 3. Test Construction of KFZA

4 Hauptkriterien	11 Skalen	26 Items	26 Einzelfragen
Arbeitsinhalte	Vielseitigkeit	Lernförderlichkeit	Können Sie bei Ihrer Arbeit neues dazulernen?
		Einsatz von Wissen	Können Sie bei Ihrer Arbeit Ihr Wissen und Können voll einsetzen?
		Abwechslungsreichtum	Bei meiner Arbeit habe ich insgesamt gesehen häufig wechselnde, unterschiedliche Arbeitsaufgaben.
	Ganzheitlichkeit	Transparenz	Bei meiner Arbeit sehe ich selber am Ergebnis, ob meine Arbeit gut war oder nicht.
		Vollständigkeit	Meine Arbeit ist so gestaltet, dass ich die Möglichkeit habe, ein vollständiges Arbeitsprodukt / eine vollständige Arbeitsaufgabe von Anfang bis Ende herzustellen.
Stressoren	Qualitative Arbeitsbelastung	Kompliziertheit	Bei dieser Arbeit gibt es Sachen, die zu kompliziert sind (z.B. aufgrund keiner oder unklarer Arbeitsbe-schreibungen oder aufgrund mangelnder Qualifizierung).
		Konzentration	Es werden zu hohe Anforderungen an meine Konzentrationsfähigkeit gestellt.
	Quantitative Arbeitsbelastung	Zeitdruck	Ich stehe häufig unter Zeitdruck.
		Arbeitsmenge	Ich habe zu viel Arbeit.
	Arbeitsunterbrechungen	Materielle Unterbrechungen	Oft stehen mir die benötigten Informationen, Materialien und Arbeitsmittel nicht zur Verfügung.
		Soziale Unterbrechungen	Ich werde bei meiner eigentlichen Arbeit immer wieder durch andere Personen unterbrochen.
	Umgebungsbelastungen	Umgebungsbedingungen	An meinem Arbeitsplatz gibt es ungünstige Umgebungs-bedingungen, wie Lärm, Klima, Staub.

Ressourcen		Raumerleben	An meinem Arbeitsplatz sind Räume und Raumausstattung ungenügend.
	Handlungsspielraum	Reihenfolge der Arbeitsschritte	Wenn Sie Ihre Tätigkeit insgesamt betrachten, inwieweit können Sie die Reihenfolge der Arbeitsschritte selbst bestimmen?
		Einfluss auf die Art der Arbeit	Wie viel Einfluss haben Sie darauf, welche Arbeit Ihnen zugeteilt wird?
		Planung und Einteilung der Arbeit	Können Sie Ihre Arbeit selbstständig planen und einteilen?
	Soziale Rückendeckung	Soziale Unterstützung durch Kollegen	Ich kann mich auf meine Kolleginnen und Kollegen verlassen, wenn es bei der Arbeit schwierig wird.
		Soziale Unterstützung durch Vorgesetzte	Ich kann mich auf meine/n direkte/n Vorgesetze/n verlassen, wenn es bei der Arbeit schwierig wird.
		Soziale Kohäsion	Man hält in der Abteilung gut zusammen.
	Zusammenarbeit	Kooperationserfordernisse	Diese Arbeit erfordert enge Zusammenarbeit mit ande-ren Kolleginnen und Kollegen in der Organisation.
		Kommunikationsmöglichkeiten	Ich kann mich während der Arbeit mit verschiedenen Kolleginnen und Kollegen über dienstliche und private Dinge unterhalten.
		Rückmeldung	Ich bekomme von Vorgesetzten und Kollegen immer Rückmeldung über die Qua-lität meiner Arbeit.
Organisationsklima	Information & Mitsprache	Information	Über wichtige Dinge und Vorgänge in unserer Organisation sind wir ausreichend informiert.
		Mitsprache	Die Leitung unserer Organi-sation ist bereit, die Ideen und Vorschläge der Beschäftigten zu berücksichtigen.
	Betriebliche Leistungen	Weiterbildungsmöglichkeiten	Unser Unternehmen bietet gute Weiterbildungsmög-lichkeiten.
		Entwicklungsmöglichkeiten	Bei uns gibt es gute Auf-stiegschancen (z.B. auch durch Erweiterung des bisherigen Tätigkeitsfeldes).

(Source: Kunz, 2015)

Nebel, Wolf and Richter (2010) then classify KFZA as a group of subjective psychological burden measurement, in which an individual evaluates the condition of the working environment he experiences. Some researchers also use and adapt KFZA for research in certain contexts, such as: working conditions in hospital (Bauer & Groneberg, 2013), work anxieties (Muschalla, Fay & Linden, 2016; Muschalla, 2017), software-ergonomic quality within computer working (Hurtienne & Prümper, 2003), Games at Work (Reinecke, 2009) and Sex role orientation and work stress (Kada, 2010). However, mostly those KFZA's studies were conducted in the context of German culture.

Method

These paper focused on the initial translation process from German language to Indonesian language. Inter rater agreement measurement was used to evaluate the level of consistency between two translators. Banerjee et al. (1999) recommended Cohen's Kappa to calculate the level of agreement between two or more raters within nominal scale. Initially, the original version of KFZA was translated using Google Translate application. Then, two translators who have Indonesian language as their mother tongue evaluated those initial translation. The first translator hold *Großes Deutsches Sprachdiplom* or a highest non-university German Language Certificate. He previously studied in Faculty of Mechanical Engineering in Aachen University of Technology since 1993 to 2002. He then decided not to completed his diploma study and returned to Indonesia. Currently he works as Indonesian-German legal translator. His German-Indonesian translation project mostly related to the laws and regulation

documents, company regulation documents, website materials, handbook, manual book and working logs, and subtitle. The second translator is a 36 years old Indonesian woman who lived in German for about 18,5 years. She got her Diploma in Information from Aachen University of Technology. Her level of German language is C2. Currently, she works as Senior User Experience Researcher in Autoscout24 GmbH and stayed in Munich. Two main data then obtained from that two translators. First, their judgment about the quality of initial translation resulted by Google Translate web application. The second is their feedback on the forward translation. This feedback might be captured some recommendation that relevant with the context of Indonesia.

These initial translate then evaluated by the two translators who provide an assessment of whether the translation is accurate or not. Cohen's kappa in SPSS statistic then run and generated two main output, namely Cross tabulation Table and Symmetric Measures Table. Cohen's kappa also measures the proportion of agreement over and above the agreement expected by chance with its range from -1 to +1 (Cohen, 1960). Cohen also suggested the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (McHugh, 2012).

Result

This paper simply evaluated the degree of agreement between Indonesian translators toward KFZA's forward translation. 26 items of KFZA initially translated by utilized Google Translate web application which can be seen in Table 3.

Table 4. Samples of KFZA's Initial Translation

Item	German Language	Indonesian Language

Item 4	Bei meiner Arbeit sehe ich selbst am Ergebnis, ob meine Arbeit gut war oder nicht.	Peluang memprediksi kualitas hasil kerja, apakah bagus atau tidak hasilnya.
Item 10	Oft stehen mir die benötigten Informationen, Materialien und Arbeitsmittel nicht zur Verfügung.	Seringkali informasi/materi/bahan/peralatan kerja yang Anda butuhkan justru tidak tersedia.
Item 16	Können Sie Ihre Arbeit selbständig planen und einteilen?	Seberapa bebas dan mandiri, Anda dapat merencanakan dan mengelola pekerjaan Anda sendiri?
Item 24	Die Leitung unseres Unternehmens ist bereit, die Ideen und Vorschläge der Beschäf-tigten zu berücksichtigen.	Seberapa banyak ide-ide dan saran dari para karyawan yang dipertimbangkan atau diperhatikan oleh pimpinan di organisasi Anda?

Table 4 shows the degree to which two translators agreed and disagreed on their judgment toward the accuracy of the KFZA's initial translations. From the 26 items of KFZA, 23 items displayed inaccurate as agreed by both translators. In addition, there was one item that displayed accurate by both the translators. Therefore, there were two items for which the two translators could not agree on the translation.

Table 5. Translator_1 * Translator_2 Cross tabulation

		Translator_2		
		Accurate	Inaccurate	Total
Translator_1	Accurate	1	1	2
	Inaccurate	1	23	24
Total		2	24	26

Table 6. Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.458	.323	2.337	.019
N of Valid Cases		26			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

Table 6 shown the Symmetric Measures which presents the Cohen's kappa (k) and statistically designed to take into account chance agreement. The coefficient of Cohen's kappa (k) is .458. This coefficient represents a weak of agreement. Furthermore, since *p*=.019 (which actually means *p<.05),* the kappa (k) coefficient is statistically significantly different from zero in the level of significant .05.

Conclusion

Evaluating the degree of agreement between the raters becomes important to address the issues of consistency and objectivity in measurement. In the context of cultural adaptation test, language adjustment to the local setting is the most critical aspect, due to the language represents the meaning behind of which want to be measured. The result of Cohen's Kappa show that the level of agreement is low. It indicates that the level of competence among the translators are not equivalent. Thus it is necessary to consider equality of expertise among translators since their recruitment. While with 23 items in the initial translation considered to be inaccurate translation, it is necessary to use more than one translator machine and tested the level of the agreement. For example, using Google translate and Online German Indonesia Translator.

The initial translation process is indeed the first step in adapting a test tool, especially in a cross-cultural context. However, the quality of the initial translation results into an effort to produce quality adaptation results early on. The next process that needs to be pursued is the reconciliation process between the translator (wild et al., 2005) and it would be better if it also involves the expert in terms of methodologies and expert in the variables being measured (Beaton et al. 2007).

REFERENCES

Apreviadezy, P. & Puspitacandri, A. (2014). Perbedaan Stres Ditinjau dari Ibu Bekerja dan Ibu Tidak Bekerja. *Jurnal Psikologi Tabularasa,* vol. 9, No. 1, pp. 58-65.

Bachroni, M. & Asnawi, S. (1999). Stres Kerja. *Buletin Psikologi,* Vol. 7, No. 2, pp. 28-39

Banerjee, M. et al. (1999). Beyond Kappa: A Review of Interrater Agreement Measures. *The Canadian Journal of Statistics,* Vol. 27, No. 1, pp. 3-23.

Bauer, J. & Groneberg, D. (2013). Perception of Stress-related Working Conditions in Hospital (iCept-study): a Comparison between Physicians and Medical Students. *Journal of Occupational Medicine and Toxicology.* Vol. 8, No. 3, pp. 1-6

Beaton, D., Bombardier, C., Guillemin, F. & Ferraz, M. B. (2007). Recommendations for the Cross-Cultural Adaptation of the DASH & Quick DASH Outcome Measures. Institute for Work & Health.

Chapman, D. W. & Carter, J. F. (1979). Translation Procedures for The Cross Cultural Use of Measurement Instrument. *Educational Evaluatin & Policy Analysis,* vol.1, No. 3, pp. 71-76.

Cohen, K. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* Vol. XX, No. 1, pp. 37-46.

Gabreya, W. (2009). IACCP, the Universe and Everything: Result of the Membership Survey, and Some Musings on Organizational Culture. *Cross-Cultural Psychology Bulletin,*vol. 43, number 1, *pp.13-22.*

Hidayati, R., Yadi, P. & Yuwono, S. (2008). Kecerdasan Emosi, Stres Kerja dan Kinerja Karyawan. *Jurnal Ilmiah Psikologi.* Vol. 2, No. 1, pp. 91-96

Hofstede, G. (n.d.). What about Indonesia? Retrieved from https://geert-hofstede.com/indonesia.html.

Howitt, D. & Cramer, D. (2011). Introduction to Research Methods in Psychology, 3rd Edition. Essex: Pearson Education Limited.

Hurtienne, J. & Prümper, J. (2003). Stress in the Office: the Influence of Software-Ergonomic Quality. In D. Harris, V. Duffy, M. Smith & C. Stephanidis (Eds.), Human-Centred Computing: Cognitive, Social, and Ergonomic Aspects (S. 63-67). Mahwah, N.J., London: Lawrence Erlbaum Associates.

Iacobucci, D. (ed.). (2001). Inter rater reliability. *Journal of Consumer Psychology's Special Issue on Methodological and Statistical Concerns of the Experimental Behavioral Researcher,* 10, pp. 71-73. Retrieved from https://pdfs.semanticscholar.org/d891/8fab-7781f11f0febdf91de297012fc0c32a8.pdf

International Test Commission. (2005, July 15). ITC Guidelines for Translating and Adapting Tests. Retrieved May 8, 2017, from https://www.intestcom.org/files/guideline_test_adaptation.pdf

Kada, O. (2010). Sex Role Orientation, Stressors and Resources in Male and Female Nurses. *Review of Psychology,* Vol. 17, No.2, pp 1-8.

Kunz, S. (2015). Gefährdungsbeurteilung psychischer Belastungen: Vergleich zweier Arbeitsanalyseverfahren am praktischen Beispiel der MOTIP DUPLI GmbH. *Bachelor Thesis.* Hochschule Furtwangen

Lubis, R. & Budiman, Z. (2014). Hubungan antara Stres Kerja dengan Kinerja Dosen di Universitas X. *Psikologia,* Vol. 9, No. 3, pp. 96-102.

Mahardayani, I. H. & Dhania, D. R. (2014). Komitmen Organisasi Karyawan *Outsourcing* Kabupaten Kudus Ditinjau dari Stres Kerja. *Humanita,* Vol. 11, No. 1, pp. 33-40.

Mahdia, A. (2014). Stres Kerja pada Tuna Netra yang Bekerja sebagai Karyawan Perusahaan Berbasis Profit di Jakarta. *Jurnal Ilmiah Psikologi,* vol. 7, no. 2, pp. 1-6

McHugh, M. L. (2012). Interrater Reliability: the Kappa Statistic. *Biochemia Medica,* 22 (3), pp. 276-282.

Muschalla, B. (2017). Is it a Case of "Work-Anxiety" When patients Report Bad Workplace Characteristics and Low Work Ability. *Journal Occupational Rehabilitation*, Vol. 27, pp. 106-114.

Muscalla, B., Fay, D. & Linden, M. (2016). Self-Reported Workplace Perception as Indicators of Work Anxieties. *Occupational Medicine*, Vol. 66, pp. 168-170.

Nebel, C., Wolf, S. & Richter, P. (2010). Instrumente und Methoden zur Messung Psychischer Belastung. Retrieved from http://www.innsicht.de/files/downloads/Nebel_Wolf_Richter_2010_Instrumente%20und%20Methoden%20zur%20Messung%20psychischer%20Belastungen.pdf

Patton, M. Q. (2015). Qualitative research & evaluation methods: integrating theory and practice, 4th Edition. London: SAGE Publication Inc.

Prasetyo, A. R. & Nurtjahjanti, H. (2012). Pengaruh Penerapan Terapi Tawa terhadap Penurunan Tingkat Stres Kerja pada Pegawai Kereta Api. *Jurnal Psikologi Undip,* vol. 11, No. 1, pp. 59-73

Prasetyo, A. R., Nurtjahjanti, H., Fauziah, N. & Kustanti, E. R. (2012). Penurunan Tingkat Stres Kerja pada Penerbang Militer Melalui Penerapan Terapi Yoga Tawa. *Jurnal Psikologi Undip,* vol. 15, No. 1, pp. 11-20.

Prumper, J., Hartmannsgruber, K. & Frese, M. (1995). KFZA. Kurz-Fragebogen zur Arbeitsanalyse. *Zeitschrift fur Arbeits- und Organisationspsychologie, 39. Jg. (N. F. 13) 3, pp. 125-132*

Safaria, T. (2011). Peran Religious Coping Sebagai Moderator dari Job Insecurity terhadap Stres Kerja pada Staf Akademik. *Humanitas,* vol. 8, No.2, pp. 155-170

Sagala, M. (2013). Efek Pelatihan Relaksasi untuk Menurunkan Stres Kerja pada Karyawan di PT Madubaru Yogyakarta. *Emphaty Jurnal Fakultas Psikologi,* Vol. 2. No. 1.

Suwartono, C. (2016). Alat Tes Psikologi Konteks Indonesia: Tantangan Psikologi di Era MEA. *Jurnal Psikologi Ulayat,* Vol. 03. No. 1, pp. 01-06.

Tinsley, H.E.A. & Weiss, D. J. (1975). Interrater Reliability and Agreement of Subjective Judgments. *Journal of Counseling Psychology,* Vol. 22, No. 4, pp. 358-376.

Triandis, H. C. (2009). The Early Stages of Cross-Cultural Psychology. *Cross-Cultural Psychology Bulletin,*vol. 43, number 1, *pp.08-12*

Widhiastuti, H. (2002). Studi Meta-Analisis tentang Hubungan Antara Stress Kerja dengan Prestasi Kerja. *Jurnal Psikologi,* No. 1, pp. 28-42.

Widodo, S. & Djalali, M. A. (2011). Stres Kerja, Kepuasan Kerja, Kesehatan Fisik, Kesalahan Kerja dan Kecelakaan Kerja Teknisi Pesawat Udara. *Jurnal Psikologi Tabularasa,* Vol. 6, No. 2, pp. 418-429

Wild, D. et al. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health.* Vol. 8, No. 2, pp. 94-104.

Wulandari, D. & Wibowo, U.D.A. (2014). Hubungan antara Konflik Peran Ganda Dengan Stres Kerja pada Perawat Wanita yang Sudah Menikah di RSUD Banyumas. *Psycho Idea,* Tahun 11, No. 1, pp. 69-78