META-ANALYSIS ON KNOWLEDGE ASSESSMENT METHODS

Joshua Chiroma Gandi[1], John Gospel Gana[2], Joseph Longshak Kosen[2], James Ali Ugbo[2], Mundi Patrick Supji[2], Peace Satzen Azi[2] and Bisan Musa[2]

[1]Department of Psychology, Nigerian Defence Academy, Kaduna

[2]Department of General and Applied Psychology, University of Jos

*** Corresponding Author's Contacts and Ids**

Phone Numbers: 08036486203, 08119150678

Email: psychoperiscope@gmail.com

Registered ORCID iD:0000-0001-6724-1296

WoScience ResearcherID:AAX-2077-2020

Loop profile:1049093

Abstract

Knowledge assessment could be a derivative of measurement theories involving metric, normative, and descriptive components. Despite soft criticisms that it only increases the precision of quantitative estimates and not a hypothesis-testing activity, meta-analysis was adopted to evaluateits assessment methodsbased on inherent empirical advantages of knowledge space theory and psychometrics. This is because even small effect(s) which do not meet criterion for arbitrary significance that may limit the remits of primary studies also have significant empirical advantage in meta-analysis. Hence,relevant candidate datawere generated from suitable studies via Google search, ScienceOpen, SciSearch, PubMed/Medline, PsycInfo, PsycARTICLES, PsycTests, and PsycEXTRA. The key pre-defined items include protocol development, information search, study eligibility, relevant studies strategy delineation, creating data collection forms, data extraction process,risk of bias assessment, individual results standardization, and overall effect(s) calculation. Initially, 439 candidate studies were identified and subjected to rigorous inclusion-exclusion processes after which only the adequately eligible 41 studies were retained for final inclusion. The result shows consistency in corroborating that knowledge structure is a domain-specific as well as a deterministic knowledge organization. It supported and further advocates for both direct and indirect assessment approaches adjudged as psychometrically appropriate. There is need for more empirical studies to further determine more superior knowledge assessment method(s) or proffer helpful way of integrating the existing ones towards maximizing impact and result.

Keywords: Assessment; Knowledge; Knowledge space; Meta-analysis; Psychometric model(s)

Introduction

The deliberate action of evaluating someone or something is an assessment or (at least) a process of assessment. Hence, assessment refers to systematic method or procedure for ascertaining the psychological characteristics of individuals or phenomena (Evers, Hagemeister, Hostmaelingen, Lindley, Muniz & Sjoberg, 2013). It could be considered, according to Gandi (2018), as a form of measurement which involves investigation and or testing in order to help predict, understand, score, diagnose, and classify the individual(s) or phenomena being assessed. Knowledge assessment is seen as a phrasal construct usually derivable from measurement theories which involve metric, normative, and descriptive components.

Measurement Theories

The measurement theories of psychometric significance are classical test theory and item response theory. Classical test theory (CTT), which originated from Edgeworth's (1888) suggestion on theory of errors, premised on the assumption that an observed score refers to the sum of true and error scores. It is most often used in empirical applications, starting from physics and astronomy to mental test scores. Taking that into consideration, in line with subsequent refinements and axiomatizations, some deliberate improvements have extended the initially formulated CTT (Borsboom, 2005). Typically, axiomatization corroborated the suggestion to decompose observed scores into their corresponding true and error scores. Thisled to formulating one of the most famous equations (i.e. Observe = True + Error) in measurement. Each item of a measurement scale using CTT reflects the construct or any levels underlying it ((Fayers & Machin, 2009). This in knowledge assessment is itself a test thatreflects underlying cognitive and emotional factors alongside their respective levels when assessing knowledge state, knowledge structure, or knowledge space. Some of the items on particular knowledge assessment measures, as might deductively or inductively be anticipated, would elicit responses that on the average are higher or lower than other items (Fayers & Machin, 2009; Doignon & Falmagne, 2015).

Item response theory (IRT), which focuses more on items of a test and not only the test itself, contributes to a great deal of hypotheses for the observed phenomena and the characteristics of persons (Reckase, 2009). Since a proper assembling of the test procedures would facilitate remedying any constraints that must be met when selecting items, Reckase (2009) hints that it is especially appreciated if the trend in test development and scoring emphasizes the item rather thanthe test as a whole. IRT models could be unidimensional or multidimensional, as the case may be. Unidimensional IRT comprises a set of models with a basic premise that the interactions of a person with the test items can be adequately represented by a mathematical expression containing a single parameter describing the characteristics of the person (Reckase, 2009). Despite its usefulness, the unidimensional IRT model is (in some cases)not without deficiency especially with reference to inadequate level of interaction in terms of complex issues. This suggests the need for more complex (i.e.

multidimensional IRT) models which accurately reflect the complexity of interactions between the respondents and the test items in such cases become necessary (Reckase, 2009). Interpreting the complexity of such interactions in knowledge assessment includes appropriate hypotheses concerning violations among individual respondents on a wide range of traits. The perceived subsets of whatever those traits may be are important for performance on specific test items (Rust & Golombok, 2009).

Multidimensional IRTportrays an idealized form of a theorized model that can only be proven false if tested using a number of observations (not just one). The idealized model, as exemplified in knowledge assessments, usually reflects mathematical expression, just as Asimov (1972) earlier described Galileo in a Church observing the swing of lamps that were hanged from the ceiling by long chains. These lamps were allowed free swings like pendulums and Galileo was reported to have recorded (using his own pulse rate) the length of time taken for each swing. From the swinging-timing observations of the lamps, a mathematical formula which idealizes reality in connection to the swing-time ratio for each pendulum was developed. The same process was used in more complex situations, such as multidimensional IRT, for idealization of reality. Multidimensional IRT models provide an approximation to the relationship between the person's characteristics and their responses to test items. In spite of the fact that it has been a complex set of models, the multidimensional IRT remains the most suitable and highly useful in simplifying reality as compared to other models that are somewhat imaginations of reality (Reckase, 2009).

Knowledge Assessment

The models upon which knowledge assessments are premised include knowledge space theory (KST) and knowledge assessment psychometrics (KAP). KST is a set-and-order theoretical framework (Doignon & Falmagne, 1999; Cosyn, Doble & Matayoshi, 2021) which proposes mathematical formalism to operationalize knowledge structures in a particular domain (Spotto, Stefanutti & Vidotto, 2010). Knowledge state, in this case, refers to collection of problems that the person is capable of solving while a knowledge structure is the collection of knowledge states containing(a) one empty set which implies that none of the problems can be solved by the individual and (b)one Q set which implies that all the problems can be solved by the individual. Knowledge space is a combinatorial structure describing the possible states of knowledge of an individual's learning. The formation of knowledge space involves modelling a domain of knowledge asa set of concepts while the feasible state of knowledge is modelled as a subset of the set containing any concepts known or knowable by individuals. Due to prerequisite relations among the set of concepts, not all the subsets are typically feasible. The knowledge space in this case, therefore, refers to a family of all feasible subsets.

Different knowledge spaces are constructed differently, using the querying experts' method, explorative data analysis method, and analysis of problem solving method. An adaptive assessment of knowledge, being the most relevant applications of KST, aims at uncovering the individual's knowledge state by presenting them with only minimum number of problems. The deterministic adaptive assessment assumes that any response behaviour is determined by the individual's knowledge state. It includes selecting a problem, from problems in some (not all) of the knowledge states, at each step of the assessment process (Degreef, Doignon, Ducamp & Falmagne, 1986). However, deterministic procedures are sometimes not realistic because they do not account for possible inconsistencies between the individual's knowledge state and any corresponding observed responses.

Falmagne and Doignon (1988b) believe that nondeterministic procedures, which include discrete and continuous, take into consideration the issue of careless errors and lucky guesses. In this wise, the discrete nondeterministic procedure premises on a deterministic procedure which provides preliminary knowledge state that seems so close to a true knowledge state (Falmagne & Doignon, 1988b). Researchers such as Anselmi, Robusto, Stefanutti and Chiusole (2016) corroborate that since presented problems update the preliminary knowledge state in accordance with observed responses, it implies that only errors can differentiate a knowledge state from other neighbouring states in the knowledge structure. According to Falmagne and Doignon (1988a), the continuous nondeterministic procedure considers a likelihood function over the knowledge structure which expresses plausibility of the knowledge states. Based on an individual's response to presented problem, the likelihood function is updated at each step of the assessment process. As soon as sufficient likelihoods are concentrated at particular knowledge state, such as uncovered knowledge state, the assessment stops. This led Hockemeyer (2002) to define assessment efficiency as the number of problems required to uncover knowledge state and also defines assessment accuracy based on the proportion of the correctly identified true knowledge states.

While using KST, Hockemeyer (2002) also found in both evaluated dimensions that continuous nondeterministic procedure has an edge over discrete nondeterministic procedure. There have been situations where knowledge structure is available while no information about the error probabilities of the problems or of the knowledge states. This could be a case of knowledge structures derived through “exerts querying” method in the knowledge domain being investigated (Dowling, 1993; Koppen, 1993), or a cognitively theorized skills that are useful/instrumental for solving problems (Duntsch, 2002; Heller, Unlu Albert, 2013). In any case, it affirms, required response data are normally collected from adequate sample to effectively estimate the adaptive assessment parameter values.

The Knowledge Assessment Psychometrics (KAP) emphasizes the models, instruments, processes and quality of its measurement. Assessment of knowledge, using psychometric method, mimics the processes of test development, administration, scoring, analysis and results interpretation. It involves measuring a single attribute with multiple component items, based on validation methods, which measure the same single attribute (i.e. latent variable). Fayers and Machin (2009) argued that when multiple items are used in assessing a variable(s), there is often a model in mind for their structural relationships. This gave rise to the fact that psychometric thinking focuses on how latent variable (in any case) manifests itself in relation to the observed variables.

In knowledge assessment psychometrics, indicator variables that reflect knowledge being sought for are, according to Borsboom (2009),enmeshedin data collected for exploration and testing to determine whether the variables fit the model. Consequent upon this, suitable training evaluation models emerged at different periods for various purposes. These models, which contribute to knowledge assessment, include Kirkpatrick model, CIRO model, Phillips ROI model, Brinkerhoff model, Kaufman's model, and Anderson model (Deller, 2020). Since psychometric theories presumed that all the items in assessment scales are indicator variables, it therefore emphasizes the construction, validation and testing of models which paves the way for appropriate assessments, analyses and interpretations (Gandi, 2018; 2020).

Psychometric tests are forms of knowledge assessments. This is why Borsboom (2009) submits that "teachers would have students tested, parents would have their children's capacities assessed, countries do test their pupils for school placements, corporate firms as well as industries test applicants and personnel for respective job positions etc". Scholars (such as Amelia, Abdallah, & Mulyadi, 2019) reasoned that the assessment types, based on their purposes, include diagnostic, formative, summative, confirmative, norm-referenced, criterion-referenced, ipsative, and portfolios assessments. The key variables in this case include intelligence, exposure/experience, socioeconomic status, physical/mental health, and others which contribute to determiningthe person's knowledge structure, knowledge state, and knowledge space (Gandi, 2018). Intelligence consists of spatial, verbal, perceptual, numerical and emotional components; exposure/experience could be sequel to learning, study, observation, or personal encounter(s); socioeconomic status defines an individual's level of being such as high, moderate or low in relation to meeting their needs; while physical/mental health explains the health situation in relation to their functioning, among other things.

In the light of the above, there seems to be a cogent need to analyze knowledge assessment methods from the perspectives of Knowledge Space Theory (KST) and Knowledge Assessment Psychometrics (KAP). This is because the Knowledge Space Theory (KST) is a preferred theoretical framework that could effectively premise such analysis (Falmagne, Albert, Doble, Eppstein & Hu, 2013; Cosyn, Doble & Matayoshi, 2021) while the Knowledge Assessment Psychometrics (KAP) is a suitable conceptual framework since such assessments seem as derivatives of measurement theories (Borsboom, 2009; Amelia, Abdallah, & Mulyadi, 2019). This may also help to bring into focus the conceptual-functional integration of knowledge structure and psychometric techniques for effective assessment that could result into a greater meaningful impact, not undermining the primary focus of meta-analyzing the overall knowledge assessment methods for greater good.

Methods

Research Design and Instruments / Techniques

The study adopted meta-analytic design to investigate knowledge assessment methods. Though seemingly nota hypothesis-testing activity (Charlton, 1996), meta-analysis significantly increases the precision of quantitative estimates and specifically has the advantage of testing the predictions of hypotheses resulting fromprimary studies (Abfalg, Bernstein & Hockley, 2017). While primary research often only culminates in the conclusion that does or does not exist, the meta-analytic study considers even small effects that do not meet arbitrary significance criterion. It is an analytical comparison where known variances of the within-group are compared to unknown variances of large-sample theory, thereby exploring suggested generalizations.

Sundry forms that constitute the study instruments include protocol development form, initial eligibility screening form, data collection and extraction forms, and checklist schedule. The techniques (i.e.methods) adopted to generate the various candidate studies as required data include Google search and accessible databases such as ScienceOpen, SciSearch, PubMed/Medline, PsycInfo, PsycARTICLES, PsycTests, and PsycEXTRA.

The statistical techniques include two one-sided tests (TOST) for two proportions' equivalence ratio test, Neyman-Pearson analysis for testing pre-specified type 1 error rate, and I-squared (I^2) statistic for heterogeneity (consistency) check. These were chosen because of their respective relevance in(a) analyzing equivalence tests, (b) clarifying the ratio of any two proportions, and (c) helping to effectively compute power and effect size (Lakens, 2017; Schirmann, 1987). Thus, these techniques are considered suitable for meta-analysis that seeks to increase the precision of quantitative estimates and the prediction of hypotheses resulting from primary studies.

Sampling of and Data Collection from Selected Candidate Studies

Firstly, an all-inclusive systematic review derived the study candidate data by selecting, evaluating, and synthesizing all available evidences as relates to knowledge assessment methods. Secondly, meta-analysis helped in combining the generated data by collating and coding them towards the most appropriate but simple testing. The meta-analytic review process was based on pre-defined participating candidate studies and the following methodical data collection steps:

Step 1: *Protocol development.* Although the study protocol does not require any formal approvalfrom an IRB or any Ethics Committee, certain aspects required an informed consent from the respective authors.Someincluded papers (considered by the authors as classified) were accorded due privacy and confidentiality, in addition to the informed consent obtained from authors.

Step 2: *Information sources and search*. The databases consist of the dates of coverage, contact with study authors for possible additional studies, dates last searched, and full electronic search strategy for database including any limits used and possibility of repetition were accorded high priority.

Step 3: *Defining eligibility criteria for the data to be included.* The adopted suitable criteria helped in defining compatible articles as well as in selecting those to be assessed for common and reliable outcomes. The generated articles were based on pre-defined keywords which includeassessment, knowledge, knowledge assessment, knowledge space, meta-analysis, psychometrics, and psychometric model(s) in relation to knowledge assessment methods.

Step 4: *Strategy delineation for identifying the relevant studies*. Study selection, i.e.the process for selecting suitable candidate studies, significantly considered the eligibility screening for both systematic review and meta-analysis in line with the research focus and relevance of study. This strategy facilitated selecting particular most appropriate candidate studies for inclusion in the analysis.

Step 5: *Creating standardized form(s)*. At this point, some independent observers who were blinded to all identifying factors have facilitated reliable data extraction. The identifying factors used in this case include: (1) authors and their institutions, (2) names of the journals, (3) sources of funding (if any), and (4) appropriate acknowledgements.

Step 6: *Data extraction and risk of bias assessment*. Respective titles and abstracts were independently screened by two independent peer-reviewers whose inputs and outputs contributed towards avoiding risk of bias in assessment and ensuring appropriate data extraction. Efforts to further eliminate bias led to implementing the extraction process by a different set of other two independent reviewers as added layer of double blind quality assurance. This ensures psychometric optimality to some extent.

Step 7: *Standardizing individual results for comparison between studies*. In order to compare various generated results, after data collection in form of suitable candidate studies, the individual results acceptable for homogeneity were standardized to something homogenous. This was complemented by the extracted mean differencesfor continuous outcomes while the odd ratios (or relative risks) were succinctly considered for binary outcomes.

Step 8: *Overall effect calculation by combining the data.* From a methodological point of view, simple arithmetic averages are not considered asbeing a significantly reliable way of comparing outcomes. Since different sample sizes have different statistical power, the weighted averages of any results with more influence than the smaller ones were used. This has implication for the ever well celebrated models of"fixed effects" and "random effects".

Results

Initial searches identified a pool of 467studies which included Google search (41studies), SciSearch (49 studies), PubMed/Medline (63 studies), PsycInfo (72 studies), PsycARTICLES (86 studies), PsycTests (44 studies), PsycEXTRA (85 studies), and manual journals (28 studies).

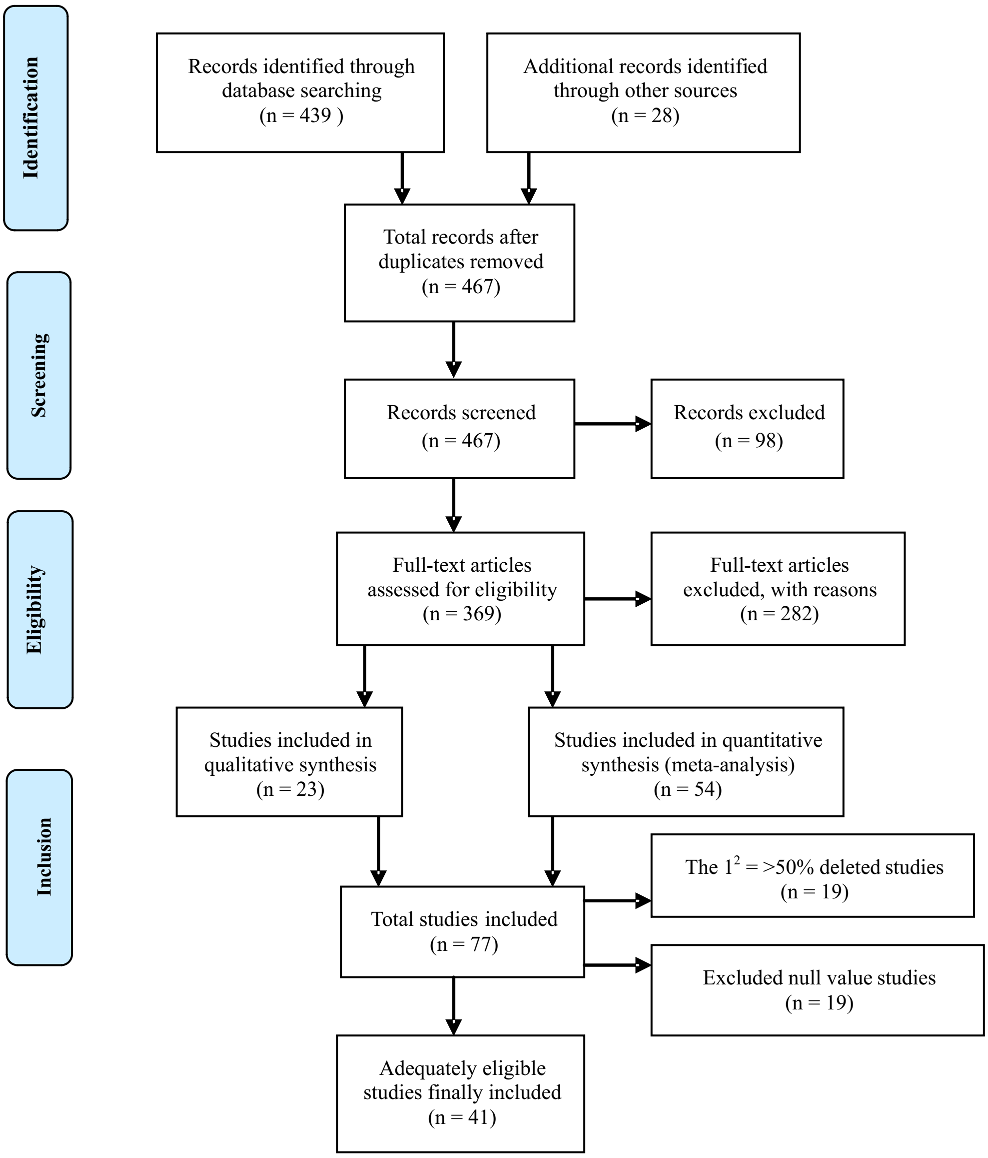

Figure 1. Article selection, screening, eligibility and inclusion flow chart

The 467 studies, following rigorous inclusion-exclusion processes (Figure 1), dropped 98 whileadditional 282 were excluded to avoid possible within-studies risk of bias.The 77 retainedstudies (23 qualitative and 54 quantitative syntheses) were subjected to heterogeneity check for consistency (I^2). The I^2 detected heterogeneity value (>50%)led to deleting 19 studies for inconsistency and another 19 for having/reflecting null values comparable to "crossing vertical line and evidently lying within 95% confidence intervals". Consequently, only 41 adequately eligible candidate studies (see Figure1) were retained for final inclusion.

Effect Size

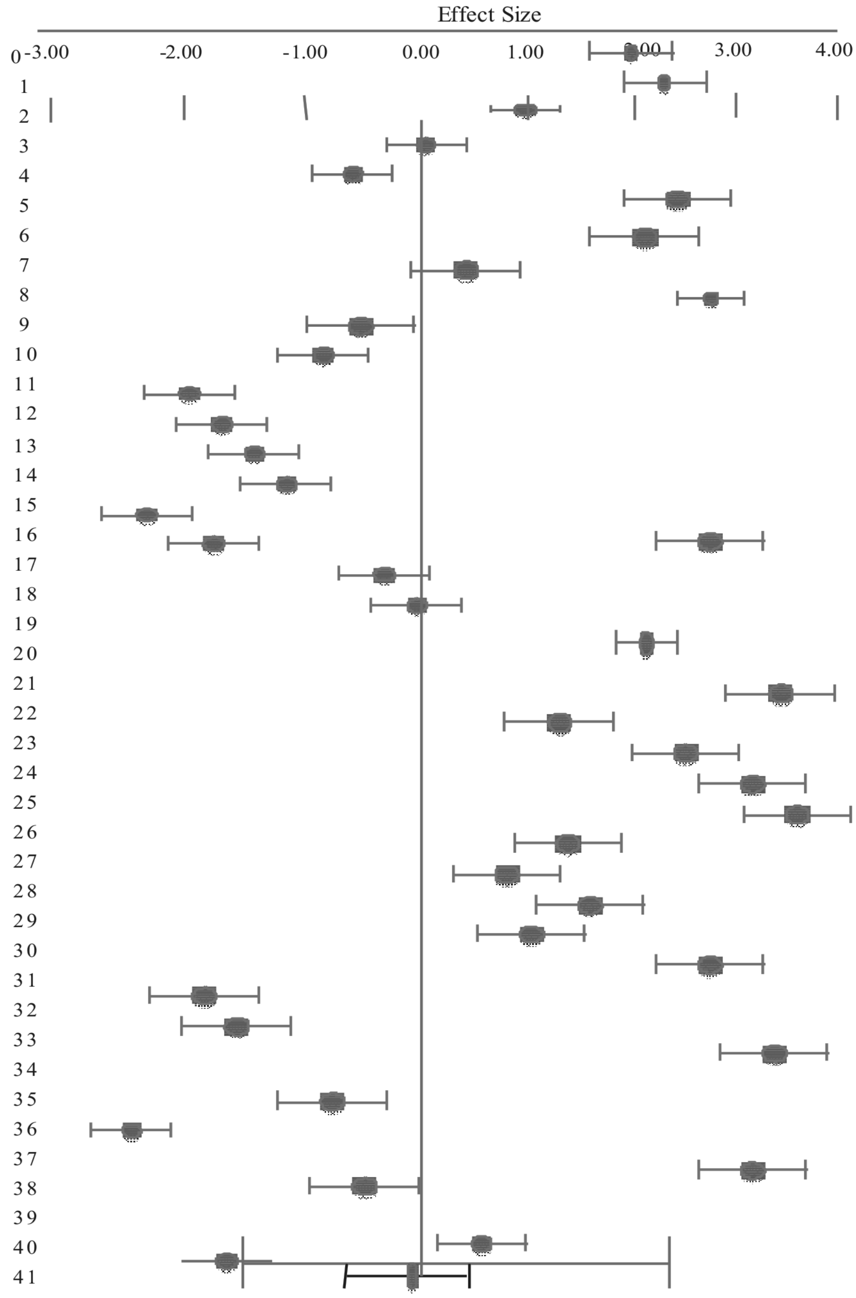

Figure 2. Forest plot showing effect size and confidence intervals

A forest plot was used (Figure 2) and it could be noted that: (a) the results of analysis have been plotted as diamond-like structures, (b) each horizontal line represents individual study with a corresponding 95% confidence interval, (c) individual candidate studies combined at the bottom and reflected their mean, and (d) horizontal points represent the limits of 95% confidence interval which applies to the combined studies as did to the individual studies. Findings from analytical comparison of within-group's known variances to the large-sample theory's unknown variances have been so revealing. It corroborates the position of Hedges, Cooper and Bushman (1992) that tests based on average effect size are usually more powerful than Stouffer test, if there is no substantial negative correlation between within-study sample size and effect size.

Discussion

The meta-analysis was performed directly on the raw difference in means because of its intuitive meaningfulness and widespread use. Since combined effects (combined p values) and tests of the weighted mean effect size are used for combined information across studies in meta-analysis (Hedges, Cooper & Bushman, 1992), a combined significance test is compared with a test based on mean effect size as the tests of null hypothesis in the study. This meta-analytic review, as succinctly summarized on flow chart (Figure 1) and forest plot (Figure 2), shows efficiency and accuracy of a continuous assessment procedure probably due to consistency check which have helped to avoid heterogeneity. This further corroborates Heller and Repitsch (2012) who observed that incorrect error probabilities (as in false negative) and incorrect prior information have the tendency to hamper efficiency and accuracy of a continuous procedure. It characteristically implies that determining knowledge truth should not necessarily be part of those favored by the initial likelihood. There are situations where knowledge structure might be available without corresponding information about the error probability of the problems, or even of the knowledge states itself. This could be the case of knowledge structures derivable from "experts querying" in a particular knowledge domain that have been under deliberate investigation (Dowling, 1993; Koppen, 1993), or the cognitively theorized skills useful/instrumental for solving problems (Duntsch, 2002; Heller, Unlu & Abert, 2013). A number of the response data collected from adequate suitable candidate studies in this case have effectively shown the adaptive assessment parameter values.

The study re-emphasized (by inference) that choosing assessment method involves ensuring the method has been designed to provide required evidence that determines the extent to which goal and outcome are achieved. Assessments as instruments are tools used for measuring knowledge, while assessments as procedures are the techniques or processes of measuring knowledge. The intended assessment goals and outcomes normally influences the choice of respective methods. This meta-analysis has further shown that there are numerous assessment methods for determining knowledge which have been broadly categorized intodirect and indirect methods.

Direct assessment methods require participants to demonstrate their knowledge, thought processes, or behavior. These are typically preferred for assessing knowledge during or after a learning situation. Typically, diagnostic, formative, confirmative, norm-referenced, criterion-referenced, ipsative, and portfolios assessments. In identifying effective communication as a goal (for instance) the direct method will involve observing and assessing participants, probably via presentation scored with a rubric. The indirect assessment methods require participants to reflect upon their knowledge, thought processes, or behavior.

These are typically preferred for assessing knowledge by way of aptitude testing approach (including diagnostic and evaluative assessments). In identifying effective oral communication as a goal (for example), the indirect method requires participants to indicate how effective they individually think they are, probably using survey-like instrument with a rating scale.

Analysis findings support the fact that deterministic models, just as a knowledge structure is to domain-specific knowledge organization, lacks realistic prediction of the person's responses to problems. The study inferred that this might have informed the introduction of a probabilistic knowledge structure, as emphasized by Anselmi, Robusto, Stefanutti and Chiusole (2016), resulting to probability distribution formula. Knowledge state (a latent construct) and response patterns (a manifest indicator of the latent construct) seems not to show perfect correspondence and this necessitates making a distinction between them. Therefore, it justifies the rationale for introducing a careless error probability and a lucky guess probability in each problem situation (Anselmi et al., 2016).

Conclusions

The study shows consistency in supporting the fact that knowledge structure is a significantly deterministic domain-specific knowledge organisation which lacks required realistic prediction of the person's responses to problems. Hence, for more efficiency and accuracy in assessing knowledge, the knowledge structure needs to be considered alongside the knowledge space and the knowledge state.Among different knowledge assessment methods, this meta-analysis supported and further advocates for diagnostic, formative, evaluative, summative and cumulative methods to be adopted in respective assessment situations. Thus, both direct and indirect methods are psychometrically appropriate as their reliable and validity have been established in the studies meta-analysed herein. This means effective monitoring and evaluation tasks, toward appropriate knowledge assessment, should be strategised in such a way that reliance on knowledge structure as domain-specific as well asknowledge space and state as deterministic factors are properly resolved.

Recommendations and Suggestions

The study has supported and further advocates for diagnostic, formative, summative, confirmative, norm-referenced, criterion-referenced, ipsative, and portfolios assessment approaches. Thesehave been the direct and indirect assessment methods adjudged psychometrically appropriate. Integrating them confirms knowledge structure being defined in the knowledge space theory as critical in effective knowledge assessments. Notwithstanding this analysis, there is need for more empirical studies to further determine more knowledge assessment method(s) or find other suitable ways of integrating the existing ones referred to in this case.

References

Abfalg, A., Bernstein, D. M., & Hockley, W. (2017). The revelation effect: A meta-analytic test of hypotheses. *Psychonomic Bulletin & Review, 24*, 1718-1741.

Amelia, N., Abdallah, A. G., & Mulyadi, Y. (2019). Meta-analysis of Student Performance Assessment Using Fuzzy Logic. *Indonesian Journal of Science & Technology, 4*(1), 21-36.

American Psychological Association (2016). Standards for Educational and Psychological Testing. Washuington DC:APA

Anselmi, P., Robusto, E., Stefanutti, L., & Chiusole, D. D. (2016). An Upgrading Procedure for Adaptive Assessment of Knowledge. *Psychometrika, 81*(2), 461-482.

Asimov, I. (1972). Asimov's new guide to science. New York: Springers.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*.UK: Cambridge University Press.

Borsboom, D. (2009). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. UK: Cambridge University Press.

Charlton, B. G. (1996). The uses and abuses of meta-analysis. *Family Practice, 12*(4), 397-401.

Cosyn, E., Uzun, H. B., Doble, C., & Matayoshi, J. (2021). A Practical Perspective on Knowledge Space Theory: ALEKS and its data. *Journal of Mathematical Psychology, 101*, 1-31.

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*, 391-418.

Cronbach, L. J., Glesser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurement: The theory of generalizability for scores and profiles*. New York: Wiley.

Degreef, E., Doignon, J.-P., Ducamp, A., Falmagne, J.-C. (1986). Languages for the assessment of knowledge. *Journal of Mathematical Psychology, 30*, 243-256.

Deller, J. (2020). Training Evaluations Models: The Complete Guide. *Kodo Survey*.

Doignon, J.-P., & Falmagne, J.-Cl. (2015). *Knowledge Spaces and Learning Spaces*. arXiv:1511.06757

Doignon, J.-P., & Falmagne, J.-Cl. (1999). *Knowledge Spaces*. Springer-Verlage. ISBN 978-3-540-64501-6

Doignon, J.-P., & Falmagne, J.-Cl. (2015).*Knowledge Space and Learning Space*. arXiv:1511.06757.

Dowling, C. E. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology, 37*, 21-48.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From Alpha to Omega: a practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412.

Duntsch, I., & Gediga, G. (2002). Skills and knowledge structures. *British Journal of Mathematical and Statistical Psychology, 48*, 9-27.

Easton, S. A., & van Laar, D. L. (2012). *User Manual for the Work-Related Quality of Life (WRQoL) Scale: A Measureof Quality of Working Life*. UK: Portsmouth University.

Edgeworth, F. Y. (1888). The Statistics of Examinations. *Journal of the Royal*

Statistical Society, 51, 598-635.

Embreston, S. E., & Reise, S. P. (2010). *Item Response Theory* (2nd ed.). New York: Routledge Academic.

Evers, A., Hagemeister, C., Hostmaelingen, A., Lindley, P., Muniz, J., & Sjoberg, A. (2013).European Federation of Psychological Associations (EFPA) Review Model for the Description and Evaluation of Psychological and Educational Tests-Version4.2.6. EFPA Board of Assessment Document 110c.

Falmagne J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology, 41*, 1-23.

Falmagne J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the state of a system. *British Journal of Mathematical and Statistical Psychology, 32*, 232-258.

Falmagne, J-C., Albert, D., Doble, C., Eppstein, D., &Hu, X. (2013). *Knowledge Spaces: Applications in Education*. Springer: New York.

Fayers, P. M., & Machin, D. (2009). *Quality of Life: The assessment, analysis and interpretation of patient-relatedoutcomes* (2nd Edn). West Sussex: Wiley.

Gandi, J. C. (2012). Empowermental partnership in coping system (EPiCS) index: measuring treatment impacts and quality of life in chronic illness. *African Journal of Applied Psychology, 6* (11&12), 25-44.

Gandi, J. C. (2018). *Development and Validation of Health Personnel Perceived Quality of Life Scale* (Unpublished doctoral thesis). University of Ibadan, Nigeria.

Gandi, J. C. (2020). Psychoperiscope. In: Wiberg M., Molenaar D., Gonzalez J., Bockenholt U., Kim J. (eds) *Quantitative Psychology*, 322, 299-318. IMPS 2019. Springer Proceedings in Mathematics and Statistics, 322, 299-318. Switzerland, Cham: Springer.

Hedges, L. V., Cooper, H., & Bushman, B. J. (1992). Testing the Null Hypothesis in Meta-Analysis: A Comparison of Combined Probability and Confidence Interval Procedures. *Psychological Bulletin, 111*(1), 188-194.Doi:10.1039 /0033-2909.111.1.188

Heller, J., Unlu, A,& Albert, D. (2013). Skills, competencies and knowledge structures. In J.-C. Falmagne, D. Albert,C. Doble, D. Epstein & X. Hu (Eds), *Knowledge Spaces: Applications in Education* (pp 229-242). Berlin, Springer.

Hockemeyer, C. (2002). A comparison of nondeterministic procedures for the adaptive assessment of knowledge. *Psychologische Beitriige, 44*, 495-503.

Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (2015). Psicometria. Porto Alegre, Artmed.

Koppen, M. (1993). Extracting human expertise for constructing knowledge spaces: An algorithm. *Journal of Mathematical Psychology, 34*, 311-331.

Lalanne, C. (2010). Applied Psychometrics with Illustrations in R (1st edn). Downloaded on 24/05/2016 from www.aliquote.org

Rao, C.R., & Sinharay, S. (2007). *Handbook of Statistics on*

***Psychometrics*, 26. Elsevier B. V.**

Reckase, M. D. (2009). *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*. NewYork:Springer.

Rust, J., & Golombok, S. (2009). *Modern Psychometrics: The Science of Psychological Assessment* (3rd edn). London: Routledge.

Spoto, A., Stefanutti, L., & Vidotto, G. (2010). Knowledge space theory, formal concept analysis, and computerized psychological assessment. *Behavior Research Methods, 42*(1), 342-350.

Stevens, S. S. (1951). Mathematics, Measurement and Psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp.1-49). New York: Wiley.

Trochim, W. M. K., & Donnelly, J. P. (2008). *The Research Methods Knowledge Base* (3rd Edition). Atomic Dog, apart of Cengage Learning.

Van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, *42*, 283-302.