

# Data Aggregation: A Proposed Psychometric IPD Meta-Analysis

Esther Kaufmann

Institute of Education, University of Zurich, Zurich, Switzerland

Email: [esther.kaufmann@gmx.ch](mailto:esther.kaufmann@gmx.ch)

**How to cite this paper:** Kaufmann, E. (2018) Data Aggregation: A Proposed Psychometric IPD Meta-Analysis. *Open Journal of Statistics*, 8, 38-48.

<https://doi.org/10.4236/ojs.2018.81004>

**Received:** December 7, 2017

**Accepted:** January 29, 2018

**Published:** February 1, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Individual participant data (IPD) meta-analysis was developed to overcome several meta-analytical pitfalls of classical meta-analysis. One advantage of classical psychometric meta-analysis over IPD meta-analysis is the corrections of the aggregated unit of studies, namely study differences, *i.e.*, artifacts, such as measurement error. Without these corrections on a study level, meta-analysts may assume moderator variables instead of artifacts between studies. The psychometric correction of the aggregation unit of individuals in IPD meta-analysis has been neglected by IPD meta-analysts thus far. In this paper, we present the adaptation of a psychometric approach for IPD meta-analysis to account for the differences in the aggregation unit of individuals to overcome differences between individuals. We introduce the reader to this approach using the aggregation of lens model studies on individual data as an example, and lay out different application possibilities for the future (e.g., big data analysis). Our suggested psychometric IPD meta-analysis supplements the meta-analysis approaches within the field and is a suitable alternative for future analysis.

## Keywords

Data Aggregation, Meta-Analysis, Bias, IPD Meta-Analysis, Psychometric Meta-Analysis, Big Data

---

## 1. Introduction

Data are the backbone of science. The technology revolution owing to the introduction of computers and Internet impacts all areas, particularly the registration and archiving of data in scientific fields. Scientific databases have increased considerably since the early days; there are more data available for analysis and interpretation. Compared to the past, repeated measurements are generally

conducted in research projects owing to registration improvements. Moreover, the goal of science is to accumulate knowledge. However, scientists require an overview of the data as a starting point. Currently, we talk of the age of “big data” because of the improvement in data gathering, registration, and archiving. Nowadays, data are often seen to have the same potential as oil did in previous years; the question that is raised is, “Do we really take advantage of our current golden oil products?” In other words, do we really know how to analyze such large datasets? Prior to the technology revolution, large datasets in social science were also analyzed, but the analysis effort involved was considerably larger than today. At the present time, one approach to analyze big data is meta-analysis. Therefore, the potential of meta-analysis approaches, particularly the so-called individual participant data (IPD) meta-analysis approach supplemented by a psychometric correction, to analyze big data, is evident.

To introduce our proposal of how to analyze data using psychometric IPD meta-analysis, we use lens model studies as an example. In our paper, we introduce psychometric IPD meta-analysis, which, to our knowledge, is unknown to the scientific community. We start to introduce the current status of meta-analysis approaches and the challenges classical meta-analysis approaches faces today. We expect IPD meta-analysis to overcome these challenges. We argue that the potential of IPD meta-analysis increases with a psychometric approach. We also introduce our suggested psychometric IPD meta-analysis by applying it to a practical example.

### 1.1. Current Meta-Analysis Approaches

Prior to the development of meta-analysis, narrative literature reviews were conducted to provide an overview of the data in a specific subject, and finally lead to a theory. The narrative review on the effect of psychotherapy by Eysenck [1] is also worth mentioning as an antecedent of the first meta-analysis method. In this review, Eysenck’s concludes that psychotherapy has no beneficial effects on patients. Glass, one of the pioneers of meta-analysis, may have been provoked by Eysenck’s conclusion. Glass also had experience as a therapist, which led him to a statistical evaluation of Eysenck’s psychotherapy review. In 1970, Glass published his meta-analysis, which aggregated the findings of 375 psychotherapy outcomes and concluded that psychotherapy does indeed work [2]. This meta-analysis is seen as one of the foundational work of modern meta-analysis approaches. As introduced by this example, the main difference between literature reviews and the further development of a meta-analysis is that literature reviews are based on studies without cumulating them. Hence, the term meta-analysis “encompasses all the methods and techniques of quantitative research synthesis” [3] and excludes traditional reviews. Since Glass [4] [5] introduced the term meta-analysis to the scientific community in his presidential speech at the American Educational Research Association, there have been numerous methodological developments [6] [7] [8]. The different meta-analysis approaches all have in

common that they aggregate data (e.g., the average judgment achievement across all judgment makers and tasks in a single study) from multiple studies.

### **1.2. Current Challenges: Heterogeneity Corrections**

With time, the focus shifted not only to the cumulation of data, but also to the explanation of the heterogeneity of data and the correction of bias. We introduce three different approaches to handle the heterogeneity within the meta-analysis results.

For example, researchers try to estimate the number of studies that were missed during the study collection for a meta-analysis. Meta-analysis is often criticized for not including all studies on a topic, which may lead to an incorrect result. This estimation is well-known as the so-called publication bias. Different types of estimates are introduced to the field and, although publication bias is often required by journal editors to have a manuscript published, we note that there is still a critical discussion on the estimation of publication bias (see [9]).

Another approach, which corrects the data heterogeneity, is the psychometric Hunter-Schmidt approach. The correction of study differences is unique to this approach. Correcting between study differences considers the fact that different studies also introduce different sources of bias such as measurement error and sampling bias, as well as the fact that data are artificially dichotomized. Since the early days of this approach, Hunter and Schmidt developed eleven so-called artifact corrections which could be applied when meta-analyzing data. For an overview of the different correction procedure, we refer to Hunter and Schmidt [7].

However, the analysis of aggregated data instead of individual-level data may introduce an ecological fallacy, because associations between two variables at the group (or ecological) level may differ from associations between analogous variables measured at the individual level (see [10], see for meta-analysis, [11], [12], p. 114). An alternative approach is to pool the individual-level data (e.g., each persons' judgment achievement in a single task) from multiple studies and analyze the pooled data directly; this is known as individual participant data (IPD) meta-analysis.

### **1.3. Individual Participant Data (IPD) Meta-Analysis**

Meta-analysis based on individual-level data has been labeled the “gold standard” of meta-analysis owing to its advantages over the classical approach [13]. Although there are several advantages to conducting an IPD meta-analysis (e.g., [11]), there are also several advantages to conducting a psychometric meta-analysis instead of a classical one, as outlined previously. Thus far, there have been no combinations thereof, namely a psychometric IPD meta-analysis. In the following method section, we suggest a psychometric IPD meta-analysis in line with the psychometric meta-analysis approach by [7] as a proposal for the missing link of IPD and psychometric meta-analysis.

## 2. Method

### 2.1. Psychometric IPD Meta-Analysis

#### 2.1.1. Database and Effect Sizes

The so-called lens model study data are ideal for our proposed psychometric IPD meta-analysis (for an overview on lens model studies, see [14] [15] [16]). Within lens model studies, the data is based on repeated judgments or measurements.

The aggregation unit in classical meta-analysis is studies, considering the different number of individuals by weighting. For example, the different number of persons within different studies is weighted by the mean of their effect size across the study. As we use lens model studies (a short introduction to lens model studies, please find below) as an example for our analysis, the effect sizes are the judgment achievements across studies. In our suggested IPD meta-analysis, the database required is individuals, for which repeated measurements are available. This assumption fits perfectly for lens model studies. Within classical meta-analysis between studies, differences based on artifacts are corrected as in the [7]. Only such an approach prevents the presumption of allocating heterogeneity based on the study and real differences. We argue that with an IPD meta-analysis based on a database of repeated measurements, individual differences are also introduced, for example measurement errors must be corrected to reveal the true individual differences. In the following, we rely on the lens model study by [17] to introduce a typical lens model study. Within these lens model studies, different teachers judge different students on their learning interests ([17]). Eighteen future education students (teachers) judged 120 students' profiles on their learning interests. Each profile includes 20 pieces of information. Each teacher's judgments are then evaluated by a test on student interest, and represented by a correlation and accuracy value. This aggregated accuracy value from the repeated judgments of one teachers is our effect size in the following. We highlight that for the following outline of data aggregation, data with repeated measurement from multiple individuals are needed. We take in the following groups of individuals from different studies. Our suggested data aggregation is also suitable for groups of individuals considering grouping factors other than studies for example schools and living regions such as Swiss cantons.

#### 2.1.2. Data Aggregation

To aggregate the introduced data, our effect size ( $r_i$ ) is a judgment achievement ( $r$ ) of teacher  $i$  and  $N_i$  is the number of judgments made by the teacher (e.g., 120 judgments on students' learning interest). Furthermore, since the sampling error is canceled out in the average correlation across individuals, we estimate the mean population correlation ( $\bar{r}$ , see Equation (1), [7]) in our data aggregation by means of the sample correlations.

$$\bar{r} = \frac{\sum [N_i r_i]}{\sum N_i} \quad (1)$$

where

$\bar{r}$  = aggregated judgment mean across individual teachers (population correlation),

$N_i$  = number of judgments made by teacher  $i$ ,

$r_i$  = judgment achievement of teacher  $i$ .

However, the sampling error due the different number of judgments made by teachers adds to the variance of correlations across persons. Therefore, the observed variance ( $\sigma_r^2$ , see Equation (2), [7], p. 100) is corrected by subtracting the sampling error variance ( $\sigma_e^2$ , 2, see Equation (3), [7], p. 100). Then, the resulting difference is the corrected variance in population correlation across persons.

$$\sigma_r^2 = \frac{\sum [N_i (r_i - \bar{r})^2]}{\sum N_i} \quad (2)$$

where

$N_i$  and  $r_i$  are as defined in Equation (1),

$\bar{r}_i$  = aggregated judgment mean of one individual teacher,

$\sigma_r^2$  = variance of the aggregated teachers' judgment achievements values (uncorrected, observed population variance).

$$\sigma_e^2 = \frac{(1 - \bar{r}^2)}{(\bar{N} - 1)} \quad (3)$$

where

$\bar{r}$  is as defined in Equation (1),

$\bar{N}$  = average number of judgments made by all teachers,

$\sigma_e^2$  = variance due to artifacts (e.g., sampling error), error variance of the aggregated teachers' judgment achievement values (error population variance).

Furthermore, the average sample size ( $\bar{N}$ ) or the average number of judgments made by a teacher has to be calculated as follows (see Equation (4), [7], p. 101):

$$\bar{N} = T/k \quad (4)$$

where

$\bar{N}$  is as defined in Equation (3),

$T$  = total number judgments across persons within one study,

$k$  = number of judgment achievements (in our case the number of teachers).

In Equation (4),  $T$  is the total number of judgments across persons, and  $k$  is the number of analyzed judgments (e.g., 370 for the number of achievement analyzed judgments across studies; [18]). Furthermore, in the meta-analysis according to Hunter and Schmidt ([7], p. 205), the credibility and confidence intervals are distinct. In contrast to the confidence intervals used, the credibility intervals do not depend on sample size; hence, the sampling error. Therefore, the credibility interval is an estimate of the range of real differences after accounting for the fact that sampling error may be due to some of the observed differences. If the lower credibility value is greater than zero, one can be confi-

dent that a relationship generalizes across persons examined in the study. As Hunter and Schmidt [7] concluded that: “credibility intervals are usually more critical and important than confidence intervals” (p. 206), we applying the 80% credibility intervals in our suggested analysis, formed by  $SD_{\rho}$  as follows (see Equation (5)):

$$\bar{\rho} = \pm 1.28 * SD_{\rho} \quad (5)$$

However, thus far, we have presented an IPD meta-analysis—applying the Hunter-Schmidt approach to individual data or taking it simply as each person is treated as a single study. Hence, in the following, we present a method of including the missing psychometric approach to IPD meta-analysis. We apply a psychometric Hunter-Schmidt approach, in which each person is again treated as a single study. However, as Hunter and Schmidt suggested up to eleven artifact corrections within the psychometric approach, we present one artifact correction, which can be taken as an example of how to apply the other artifacts to our suggested psychometric IPD approach. Within our example study, [17] reported for each teacher, retest-reliability values range from 0.2 to 0.99. We use these values for our psychometric IPD meta-analysis. The fully corrected mean correlation ( $\bar{R}$ ) or the fully corrected mean of teacher judgment achievement in a psychometric IPD meta-analysis is the corrected mean correlation in a classical IPD meta-analysis ( $\bar{r}$ , see Equation (1)) divided by the attenuation factor, as shown in Equation (6):

$$\bar{R} = Ave(\rho) = \frac{\bar{r}}{A} \quad (6)$$

where

$\bar{r}$  is as defined in Equation (1),

$A$  = Attenuation factor (artifacts, e.g., measurement error),

$\bar{R} = Ave(\rho)$  = fully corrected mean of teachers’ aggregated judgment achievement values (*i.e.*, population correlation)

In the next step, we estimate the variance in the corrected correlation across persons owing to artifact variance such as measurement error introduced by a single person. Therefore, we compute the sum of the squared coefficient of variation ( $V$ ) across the attenuation factors (see Equation (7)):

$$V = \frac{SD(a)^2}{Ave(a)^2} + \frac{SD(b)^2}{Ave(b)^2} + \dots \quad (7)$$

where

$V$  = variation across the attenuation factors,

$a$  = artifacts (e.g., measurement error) of teacher  $a$ ’s aggregated judgment achievement,

$b$  = artifacts (e.g., measurement error) of teacher  $b$ ’s aggregated judgment achievement.

Furthermore, we estimate the variance ( $S$ ) in the correlations corrected across persons, accounted for by the variation in artifacts as a product (see Equation

(8)).

$$S^2 = \bar{R}^2 \bar{A}^2 V \quad (8)$$

where

$S$  = variance of the corrected teachers' judgment achievement for all teachers,  $R$  and  $A$  are as defined in Equation 6, and  $V$  is defined Equation (7).

The unexplained residual variance ( $S_1^2$ ) in the corrected correlation across persons is calculated (see Equation (9)):

$$S_1^2 = \bar{R}^2 - S^2 \quad (9)$$

$S_1^2$  = unexplained residual variance for the other parameters, see Equations (6) ( $\bar{R}$ ) and (8) ( $S$ ).

Consequently, the fully corrected variance ( $Var(\rho_j)$ ) across persons in our proposed psychometric IPD meta-analysis is as follows (see Equation (10)):

$$Var(\rho_j) = \frac{S_1^2 - S^2}{\bar{A}^2} \quad (10)$$

where

$Var(\rho_j)$  = fully corrected variance of the aggregated teachers' judgment achievement across teachers,

$S_1^2$  is as defined in Equation (9),

$S$  is as defined in Equation (8),

$\bar{A}$  is as defined in Equation (6).

Finally, to estimate if the differences between individuals are really differences and do not rely on artifacts, the 75% rules are estimated in line with Hunter and Schmidt [7]. Hunter and Schmidt suggested subtracting the variation due to sampling error from the total variation. If artifacts remove approximately 75% of the overall variation, they conclude that the effect sizes are homogeneous. If the value is below 75%, then the lack of homogeneity of a single effect sizes is indicated and a search for moderating variables is conducted.

### 3. Results of the Psychometric IPD Meta-Analysis

#### 3.1. Database

For simplification, we apply the introduced psychometric IPD meta-analysis to the data by [17]. We supplement this data base with a second study by Levi [19]. Both studies are lens model studies. These studies are ideal for our outlined data analysis. The lens model characteristics of both studies are included in **Table 1**; for details, we refer to the original studies.

In our analysis, we consider the measurement error and sampling bias and ignore any additional artifacts. It is important to note that our suggested approach is also applicable to additional artifacts, but these are ignored in the following example for simplification. The complete databases required for our analysis is available in **Table A1** in the **Appendix**. For our analysis, we require a judgment achievement (correlation) value, the number of judgments each

**Table 1.** Summary of the specific lens model characteristics of studies included in our analysis.

Study	Lens Model Characteristics				
	Number and sample of judges	Number of judgments	Cues	Judgment task	Criterion
1. Athanasou & Cooksey [15]	18 technical and further education students	100 student profiles, 20 replications	20	Deciding whether students are interested in learning	Actual level of students' interest test
2. Levi [17]	9 nuclear medicine physicians	280 patient cases, 60 replications	5	Assess probability of significant coronary artery disease	Coronary angiography

participant made, for each participant. Thus far, it is possible to conduct a classical IPD meta-analysis with this data. Additional data is required for a psychometric IPD meta-analysis. We require two reliability values. First, we need the reliability value of the judgment made by judgment- and decision-makers. Second, we need the reliability value of the criterion values. We note that there are two different criteria (interest test/coronary angiography). The retest value of the test is taken from the literature [20]. We assume the value of the coronary angiography is quite high, leading us to use a value of .99 in our example. We choose one type of reliability values, namely retest-reliability values.

### 3.2. Analyses

We ran the simulation using all the information required for a psychometric IPD meta-analysis as outlined in our method section. We emphasize that we used the so-called Hunter-Schmidt psychometric meta-analysis program [7]; however, instead of using studies as aggregation levels, we considered single individuals as aggregation levels. The results are listed in **Table 2**.

To interpret the results of our suggested psychometric IPD meta-analysis, we argue that the single judgment achievement across these two tasks is moderate (0.52), and there is only a small heterogeneity. Owing to the 75% rule, no search for additional moderator variables is indicated. Hence, we conclude that individual variance is as a result of uncorrected artifacts in classical meta-analysis approaches, which leads to an overestimation of study variation based on uncorrected individual differences. Hence, we realize the need to focus first on the individual level and obtain an accurate data, before any aggregation or further correction should take place. This is to ensure that the data variance at the study level is not overestimated.

## 4. Conclusions and Outlook

In this paper, we introduced a psychometric IPD meta-analysis, adapting the psychometric Hunter-Schmidt approach instead of the aggregation-unit of studies to the aggregation-units of persons. To apply our suggested IPD meta-analysis successfully, we require a special data type. Our data example is based



**Table 2.** Result of our psychometric IPD meta-analysis.

Number of correlations (k)	Total sample size (n)	m	SD	80 CI	%
27	4320	0.52	0.03	0.49 - 0.56	97

m = Mean true score correlation; SD = Standard deviation of true score correlation; 80 CI = 80% Credibility Interval (10% CI, 90% CI); % = Percentage variation in observed correlation attributable to all artifacts (75%).

on a repeated measurement by a single individual, typically introduced by lens model studies. However, we note that there are various possibilities for applying our proposal to datasets outside the lens model approach. We note that ambulatory assessment data (see [21]), particularly studies applying the so-called experience sampling approach, may be a suitable future application of psychometric IPD meta-analysis.

Owing to recent improvements in registration and archiving, we can expect to see more repeated-measures studies in the future. In particular, big data involves different sources of data. Hence, our suggested IPD meta-analysis approach also has the potential for future big data analysis considering different data sources such as study differences. In future, additional developed add-ons to the classical meta-analysis approach as a cumulative meta-analysis approach could be adapted for IPD meta-analysis. We see considerable potential in transferring the aggregation unit from the study unit to the individual level. However, we note that in future, comparisons of different aggregation units will be required to increase the accuracy of data aggregation.

To summarize, our proposed method of data aggregation is not limited only to future meta-analysis, but could be applied to overall data aggregation, provided individual data and multiple measure points are available. Hence, in decision-making areas where single individuals often make multiple judgments, such an approach could be applied for data aggregation. Future research on the evaluation of the classical data aggregation approach and our proposed aggregation approach will show the current potential of our suggestion. We note that our introduced aggregation analysis is time-consuming and also requires considerable additional data. However, we believe that technology-based developments will overcome this challenge successfully, and therefore support the adoption of our proposed analysis approach in the future.

## References

- [1] Eysenck, H.J. (1952) The Effects of Psychotherapy: An Evaluation. *Journal of Consulting Psychology*, **16**, 319-324. <https://doi.org/10.1037/h0063633>
- [2] Smith, M.L. and Glass, G.V. (1977) Meta-Analysis of Psychotherapy Outcome Studies. *American Psychologist*, **32**, 752-760. <https://doi.org/10.1037/0003-066X.32.9.752>
- [3] Lipsey, M.W. and Wilson, D.B. (1993) The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from Meta-Analysis. *American Psychologist*, **48**, 1181-1209. <https://doi.org/10.1037/0003-066X.48.12.1181>

- [4] Glass, G.V. (1976) Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, **5**, 3-8. <https://doi.org/10.3102/0013189X005010003>
- [5] Glass, G.V. (2016) One Hundred Years of Research: Prudent Aspirations. *Educational Researcher*, **45**, 69-72. <https://doi.org/10.3102/0013189X16639026>
- [6] Rosenthal, R. and DiMatteo, M.R. (2001) Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology*, **52**, 59-82. <https://doi.org/10.1146/annurev.psych.52.1.59>
- [7] Schmidt, F.L. and Hunter, J.E. (2014) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage, Los Angeles.
- [8] Shadish, W.R. (2015) Introduction to the Special Issue on the Origins of Modern Meta-Analysis. *Research Synthesis Methods*, **6**, 219-220. <https://doi.org/10.1002/jrsm.1148>
- [9] Rothstein, H.R. (2008) Publication Bias as a Threat to the Validity of Meta-Analytic Results. *Journal of Experimental Criminology*, **4**, 61-81. <https://doi.org/10.1007/s11292-007-9046-9>
- [10] Robinson, W.S. (1950) Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, **15**, 351-357. <https://doi.org/10.1093/ije/dyn357>
- [11] Kaufmann, E., Reips, U.-D. and Maag Merki, K. (2016) Avoiding Methodological Biases in Meta-Analysis: Use of Online Versus Offline Individual Participant Data (IPD) in Educational Psychology. *Zeitschrift für Psychologie*, **224**, 157-167. <https://doi.org/10.1027/2151-2604/a000251>
- [12] Viechtbauer, W. (2007) Accounting for Heterogeneity via Random-Effects Models and Moderator Analyses in Meta-Analysis. *Zeitschrift für Psychologie*, **215**, 104-121. <https://doi.org/10.1027/0044-3409.215.2.104>
- [13] Chalmers, I. (1993) The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care. *Annals of the New York Academy of Sciences*, **703**, 156-163. <https://doi.org/10.1111/j.1749-6632.1993.tb26345.x>
- [14] Hammond, K.R. and Stewart, T.R. (2001) *The Essential Brunswik: Beginnings, Explications, Applications*. University Press, Oxford.
- [15] Karelaia, N. and Hogarth, R. (2008) Determinants of Linear Judgment: A Meta-Analysis of Lens Studies. *Psychological Bulletin*, **134**, 404-426. <https://doi.org/10.1037/0033-2909.134.3.404>
- [16] Kaufmann, E., Reips, U.-D. and Wittmann, W.W. (2013) A Critical Meta-Analysis of Lens Model Studies in Human Judgment and Decision-Making. *PLoS ONE*, **8**, e83528. <https://doi.org/10.1371/journal.pone.0083528>
- [17] Athanasou, J.A. and Cooksey, R.W. (2001) Judgment of Factors Influencing Interest: An Australian Study. *Journal of Vocational Education Research*, **26**, 77-96. <https://doi.org/10.5328/JVER26.1.77>
- [18] Kaufmann, E. (2010) *Flesh on the Bones: A Critical Meta-Analytical Perspective of Achievement Lens Studies*. Doctoral Dissertation, University of Mannheim, MADOC, Mannheim. <http://madoc.bib.uni-mannheim.de/madoc/volltexte/2010/2892>
- [19] Levi, K. (1989) Expert Systems Should Be More Accurate than Human Experts: Evaluation Procedures from Human Judgment and Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**, 647-657. <https://doi.org/10.1109/21.31070>
- [20] Athanasou, J.A. (2006) A Career Interest Test: A Brief, Standardised Assessment of Interests for Use in Educational and Vocational Guidance. *Revista Española de*

*Orientación y Psicopedagogía*, 17, 5-17.

- [21] Kaufmann, E. (2009) Ambulatory Assessment: A Modern Version of Brunswik's Representative Design Approach. *The Brunswik Society Newsletter*, 24, 21-22.

## Appendix

**Table A1.** Consideration of data in our meta-analysis example.

Study	Number of individuals	Number of judgments	Judgment achievement	Retest-Reliability <sup>1</sup>	Retest-Reliability <sup>2</sup>
Athanasou & Cooksey [15]	1	100	0.32	0.995	0.85
	2	100	0.563	0.346	0.85
	3	100	0.005	0.974	0.85
	4	100	0.294	0.377	0.85
	5	100	0.474	0.023	0.85
	6	100	0.435	10.00	0.85
	7	100	0.404	0.861	0.85
	8	100	0.317	0.266	0.85
	9	100	0.554	0.550	0.85
	10	100	0.290	0.807	0.85
	11	100	0.238	0.634	0.85
	12	100	0.356	0.194	0.85
	13	100	0.028	0.227	0.85
	14	100	0.185	0.322	0.85
	15	100	0.232	0.723	0.85
	16	100	0.223	0.704	0.85
	17	100	0.219	0.856	0.85
	18	100	0.377	0.913	0.85
Levi [17]	1	280	0.521	0.856	0.99
	2	280	0.503	0.903	0.99
	3	280	0.510	0.929	0.99
	4	280	0.459	0.852	0.99
	5	280	0.483	0.918	0.99
	6	280	0.458	0.92	0.99
	7	280	0.429	0.934	0.99
	8	280	0.383	0.732	0.99
	9	280	0.459	0.854	0.99

<sup>1</sup> = Reliability values of judgments; <sup>2</sup> = Reliability values of evaluation criteria.