TUS

How can we increase the credibility of technology education research?

Dr Jeffrey Buckley, PhD

Technological University of the Shannon: Midlands Midwest, Ireland
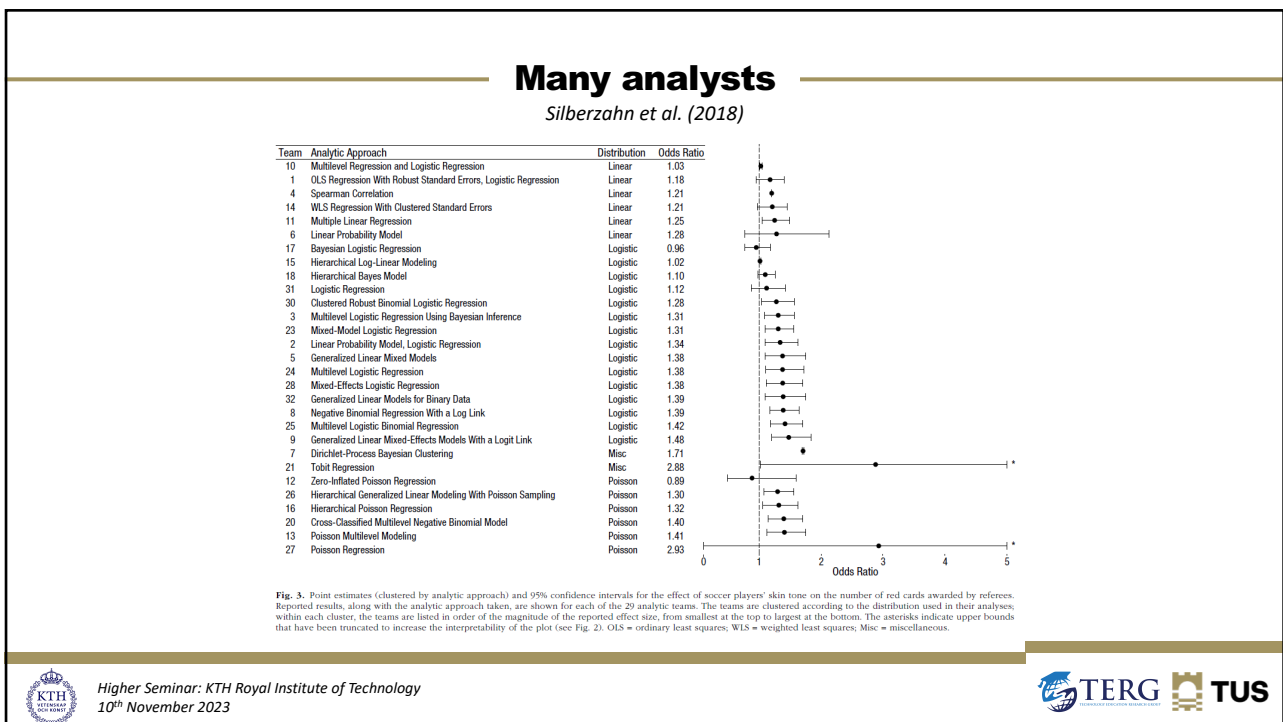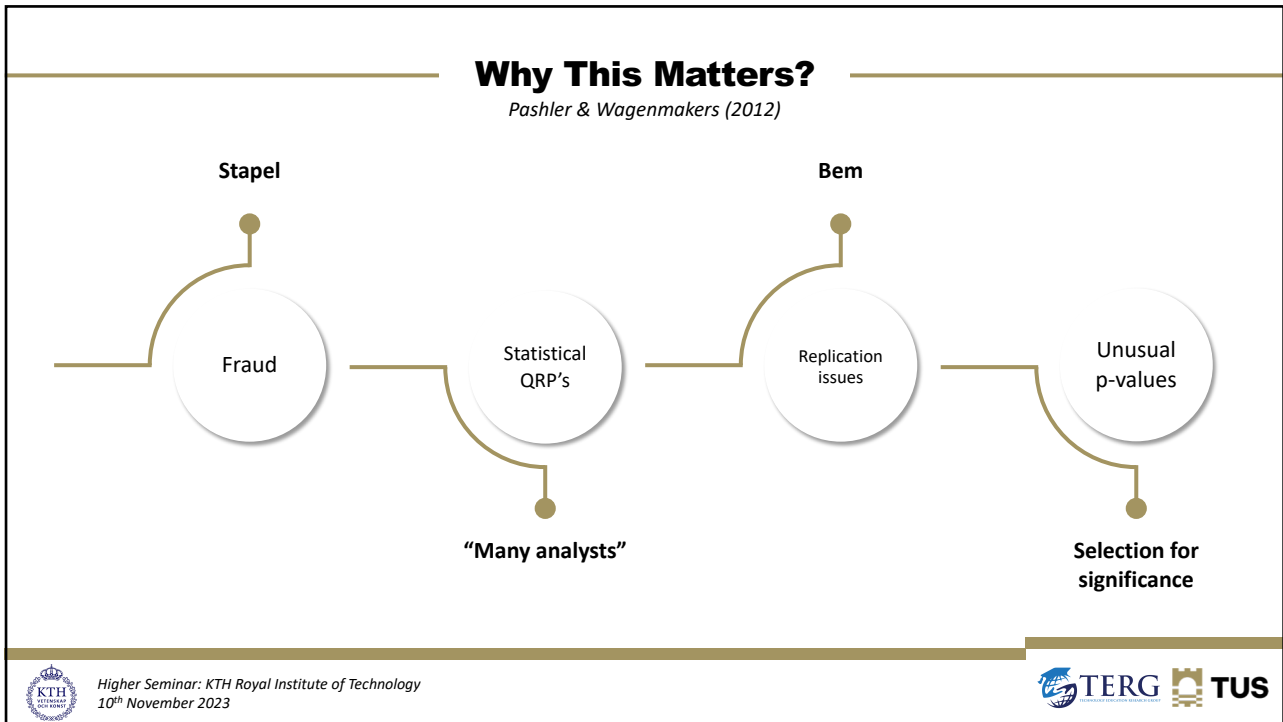

An Important Philosophy

> " We should be the most critical as we are
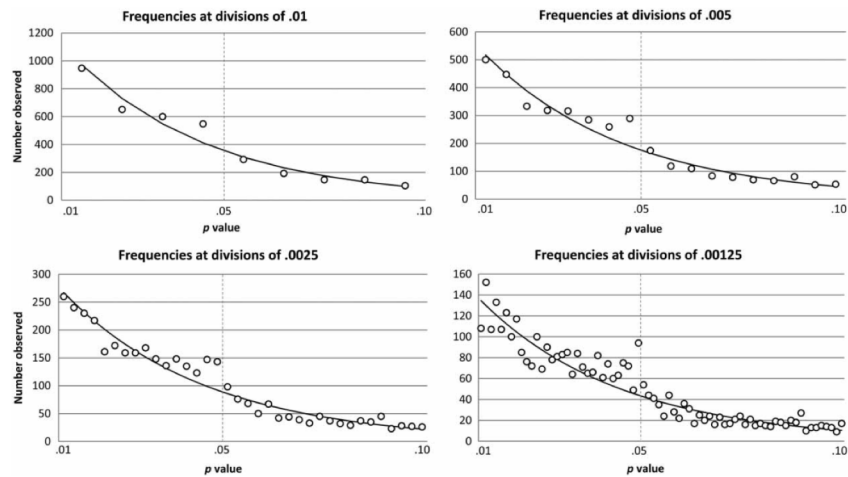> amongst friends (Spendlove, 2023). "

Why This Matters?

Pashler & Wagenmakers (2012)

Stapel

Bem

Fraud

Statistical QRP's

Replication issues

Unusual p-values

"Many analysts"

Selection for significance

Many analysts

Silberzahn et al. (2018)

Team	Analytic Approach	Distribution	Odds Ratio
10	Multilevel Regression and Logistic Regression	Linear	1.03
1	OLS Regression With Robust Standard Errors, Logistic Regression	Linear	1.18
4	Spearman Correlation	Linear	1.21
14	WLS Regression With Clustered Standard Errors	Linear	1.21
11	Multiple Linear Regression	Linear	1.25
6	Linear Probability Model	Linear	1.28
17	Bayesian Logistic Regression	Logistic	0.96
15	Hierarchical Log-Linear Modeling	Logistic	1.02
18	Hierarchical Bayes Model	Logistic	1.10
31	Logistic Regression	Logistic	1.12
30	Clustered Robust Binomial Logistic Regression	Logistic	1.28
3	Multilevel Logistic Regression Using Bayesian Inference	Logistic	1.31
23	Mixed-Model Logistic Regression	Logistic	1.31
2	Linear Probability Model, Logistic Regression	Logistic	1.34
5	Generalized Linear Mixed Models	Logistic	1.38
24	Multilevel Logistic Regression	Logistic	1.38
28	Mixed-Effects Logistic Regression	Logistic	1.38
32	Generalized Linear Models for Binary Data	Logistic	1.39
8	Negative Binomial Regression With a Log Link	Logistic	1.39
25	Multilevel Logistic Binomial Regression	Logistic	1.42
9	Generalized Linear Mixed-Effects Models With a Logit Link	Logistic	1.48
7	Dirichlet-Process Bayesian Clustering	Misc	1.71
21	Tobit Regression	Misc	2.88
12	Zero-Inflated Poisson Regression	Poisson	0.89
26	Hierarchical Generalized Linear Modeling With Poisson Sampling	Poisson	1.30
16	Hierarchical Poisson Regression	Poisson	1.32
20	Cross-Classified Multilevel Negative Binomial Model	Poisson	1.40
13	Poisson Multilevel Modeling	Poisson	1.41
27	Poisson Regression	Poisson	2.93

Fig. 3. Point estimates (clustered by analytic approach) and 95% confidence intervals for the effect of soccer players' skin tone on the number of red cards awarded by referees. Reported results, along with the analytic approach taken, are shown for each of the 29 analytic teams. The teams are clustered according to the distribution used in their analyses; within each cluster, the teams are listed in order of the magnitude of the reported effect size, from smallest at the top to largest at the bottom. The asterisks indicate upper bounds that have been truncated to increase the interpretability of the plot (see Fig. 2). OLS = ordinary least squares; WLS = weighted least squares; Misc = miscellaneous.

Selection for significance

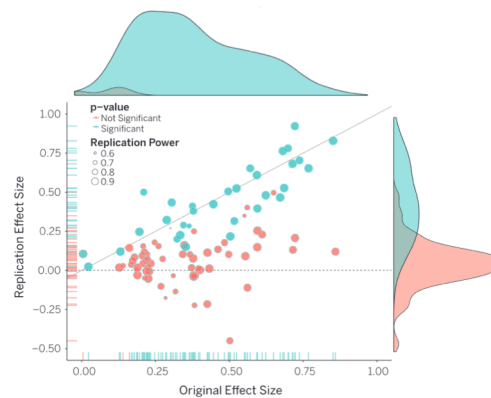*Masicampo & Lalande (2012)*

Higher Seminar: KTH Royal Institute of Technology
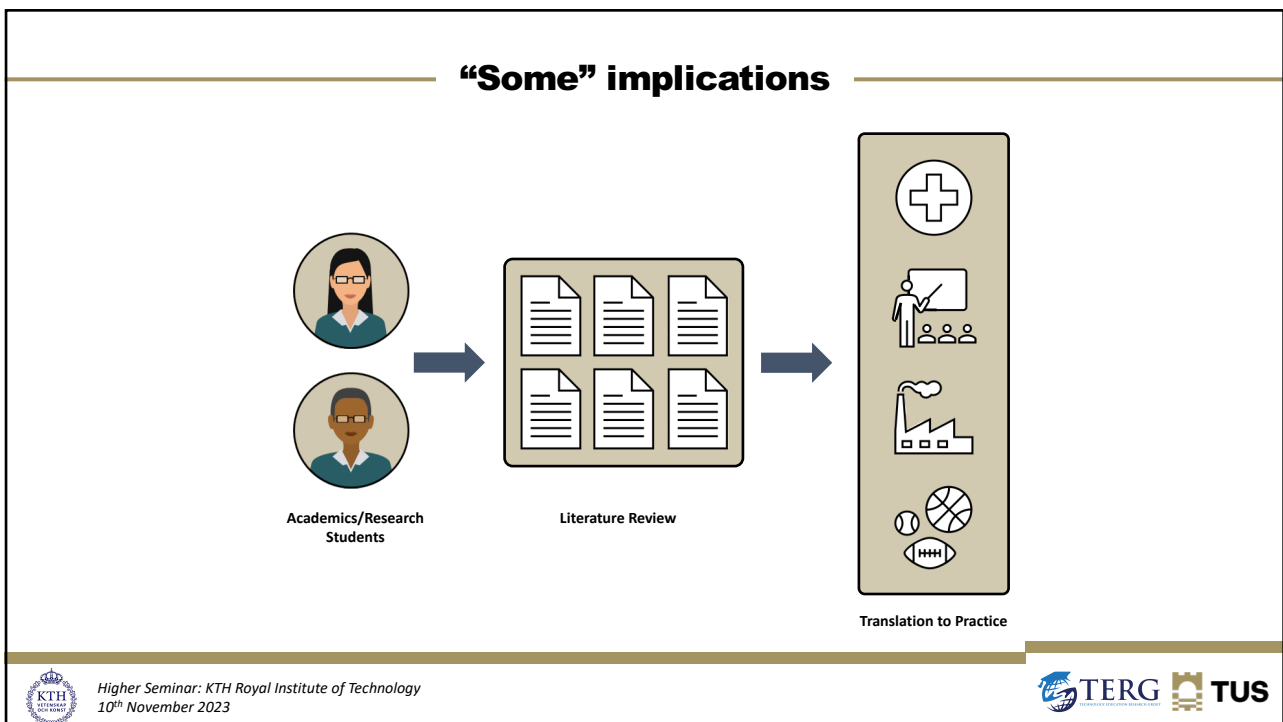*10th November 2023*

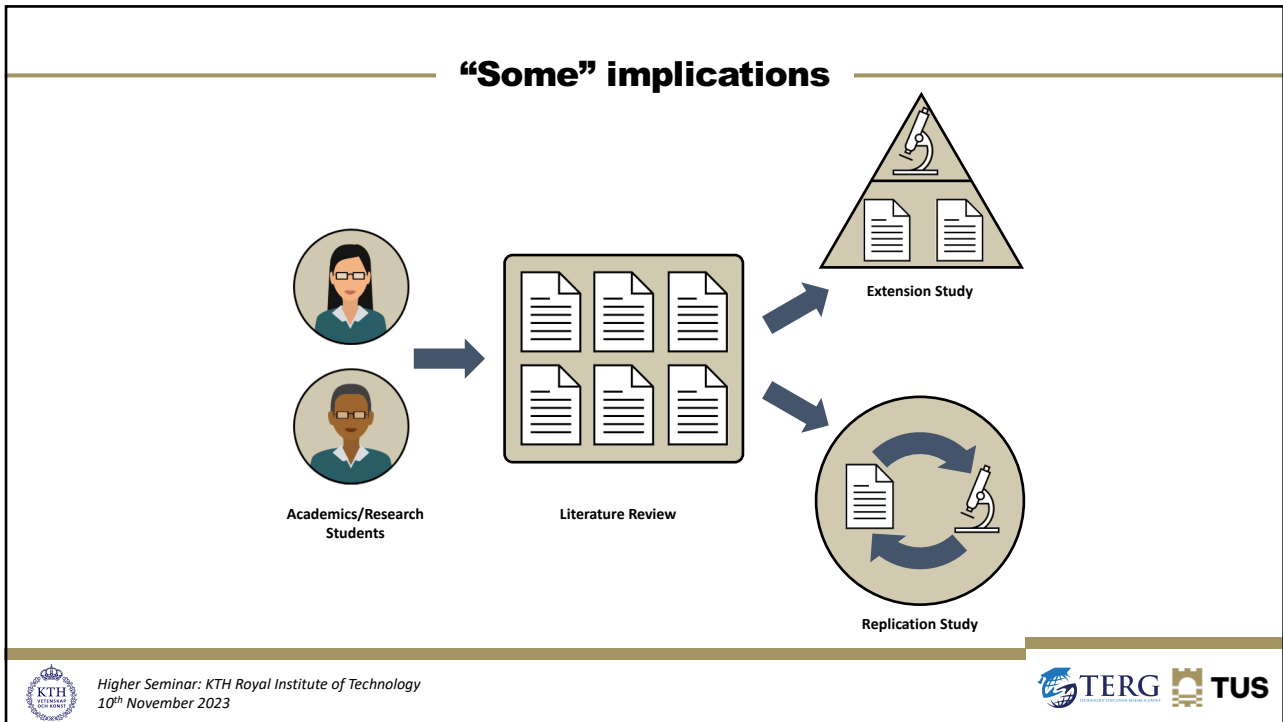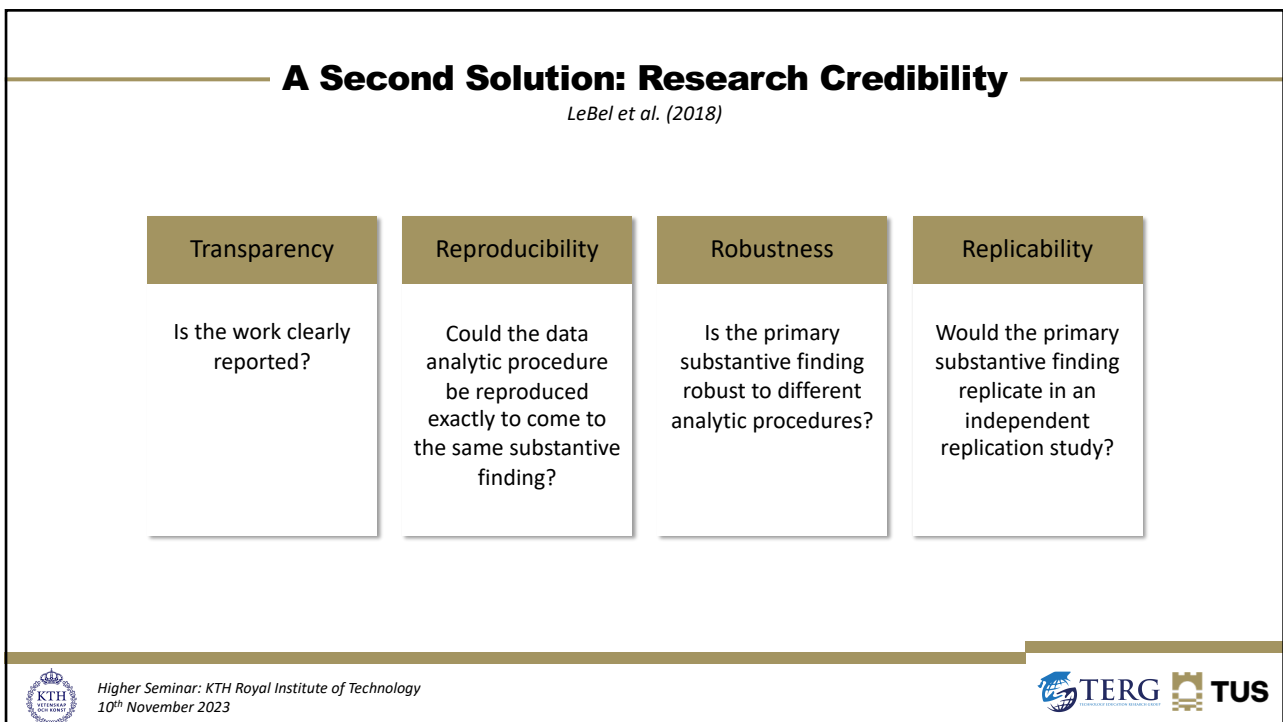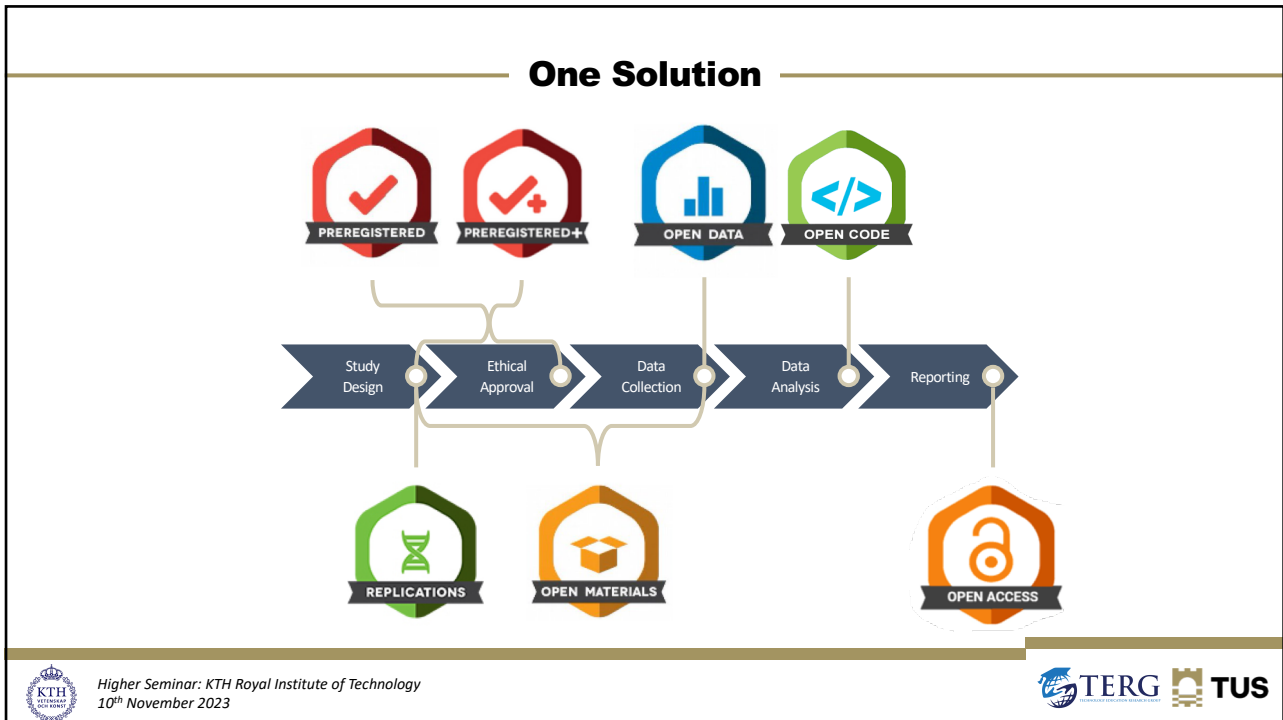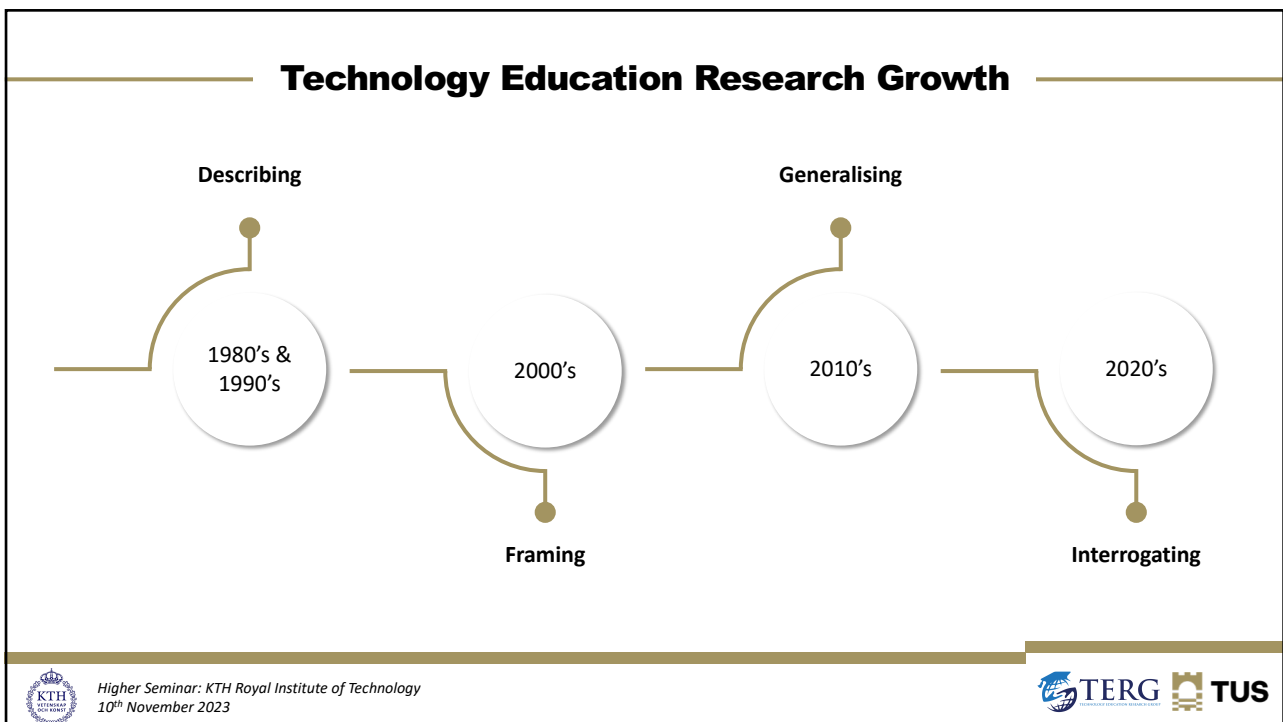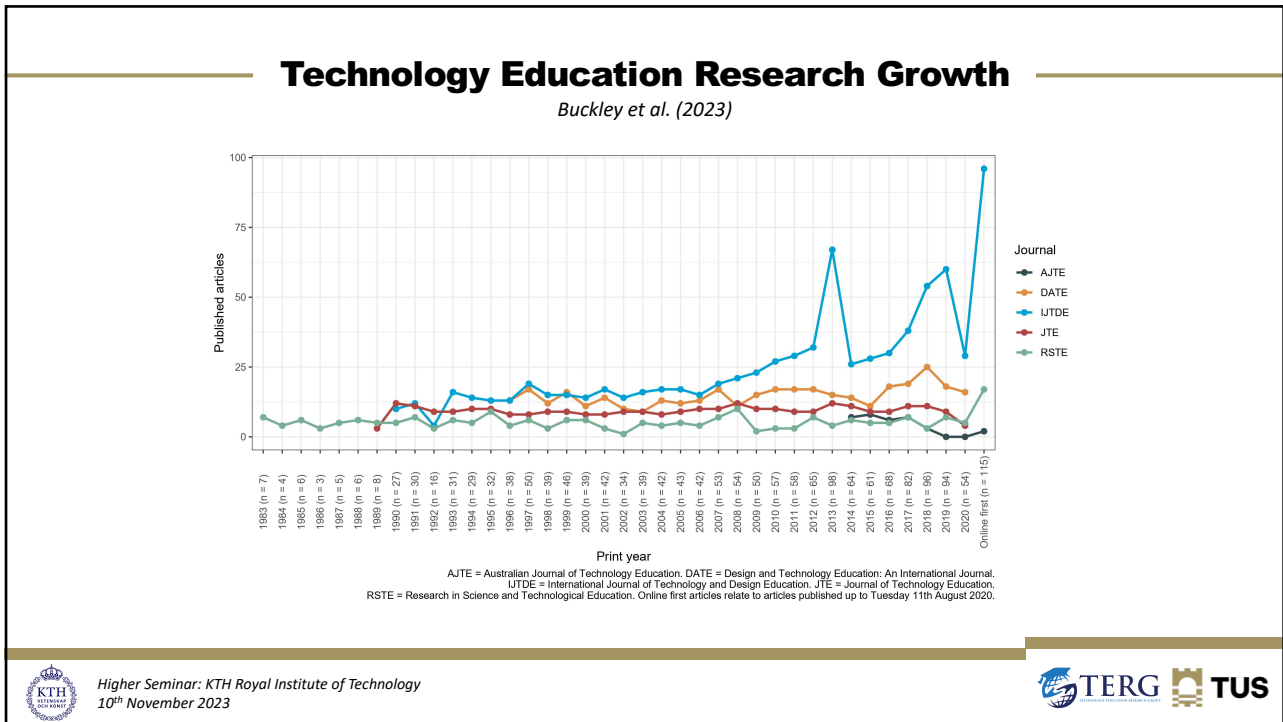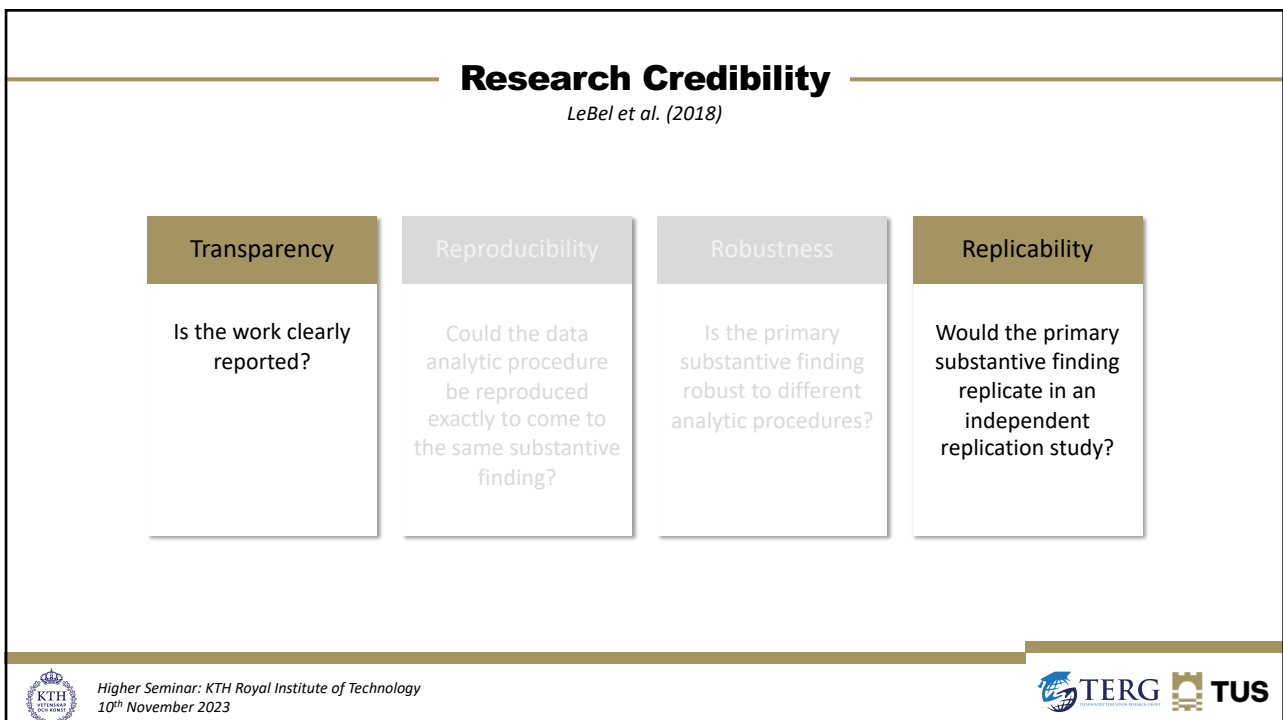# Open Science Collaboration

OSC (2015)

Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.
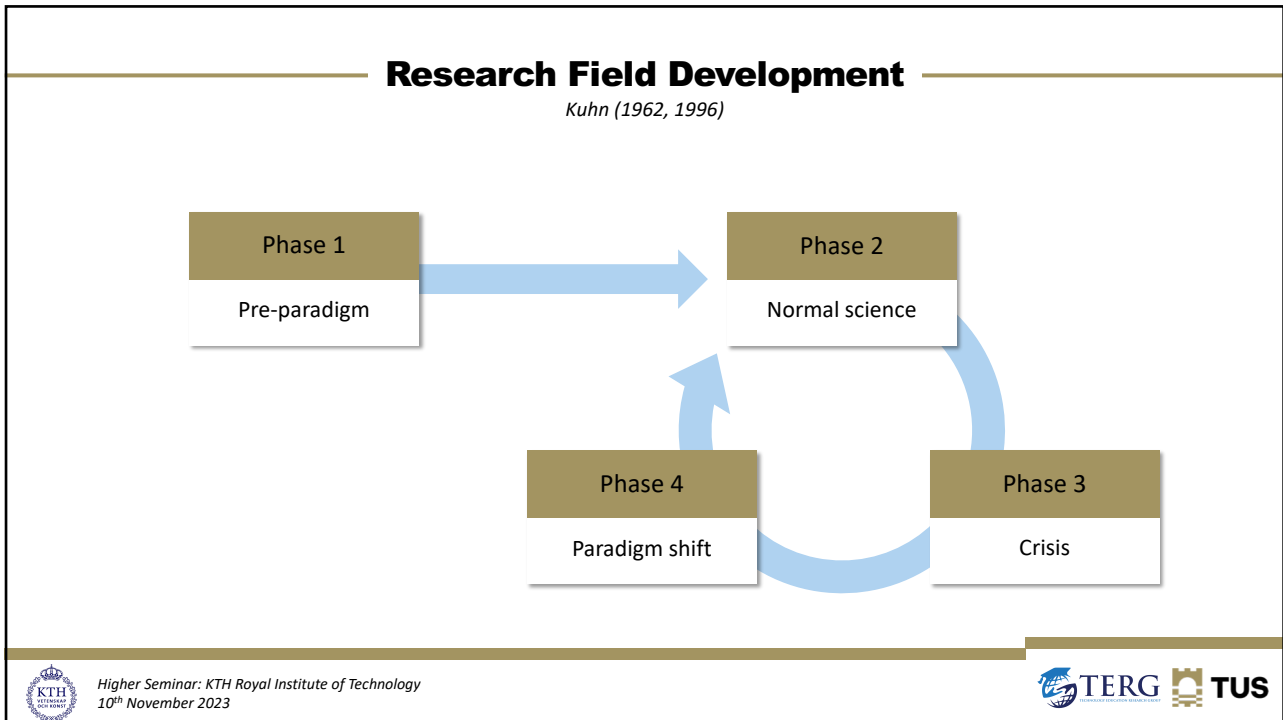
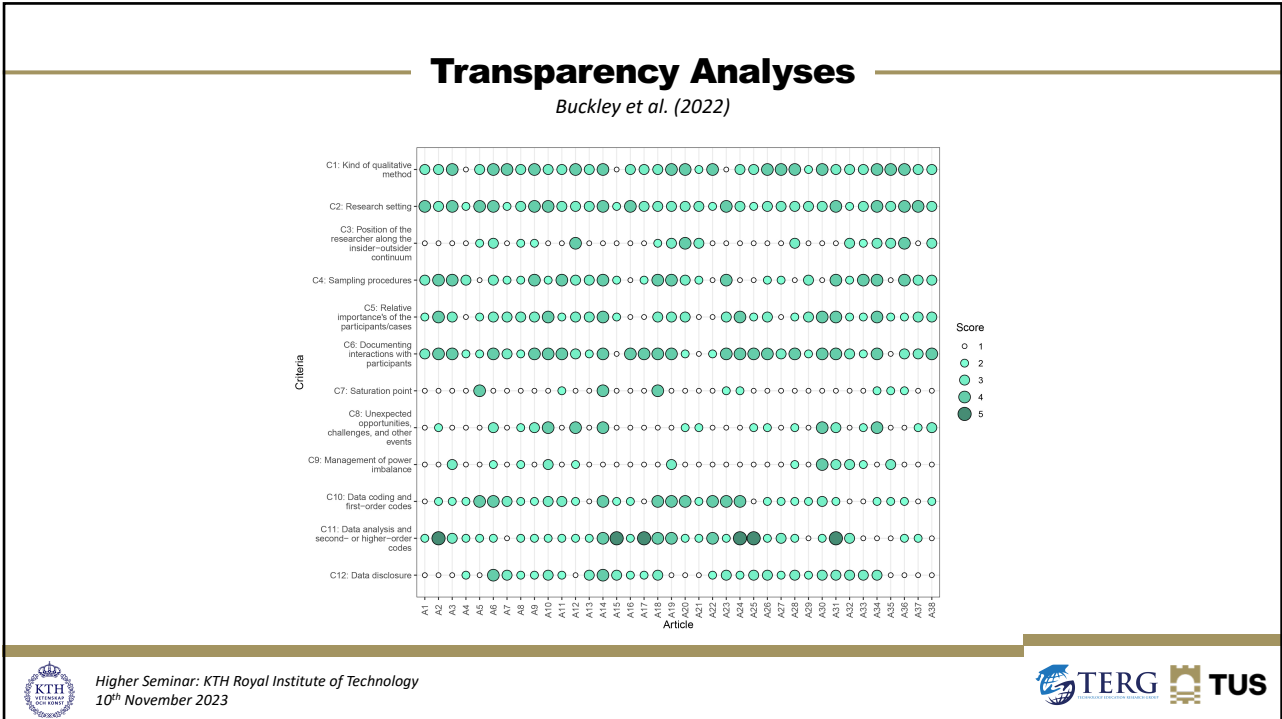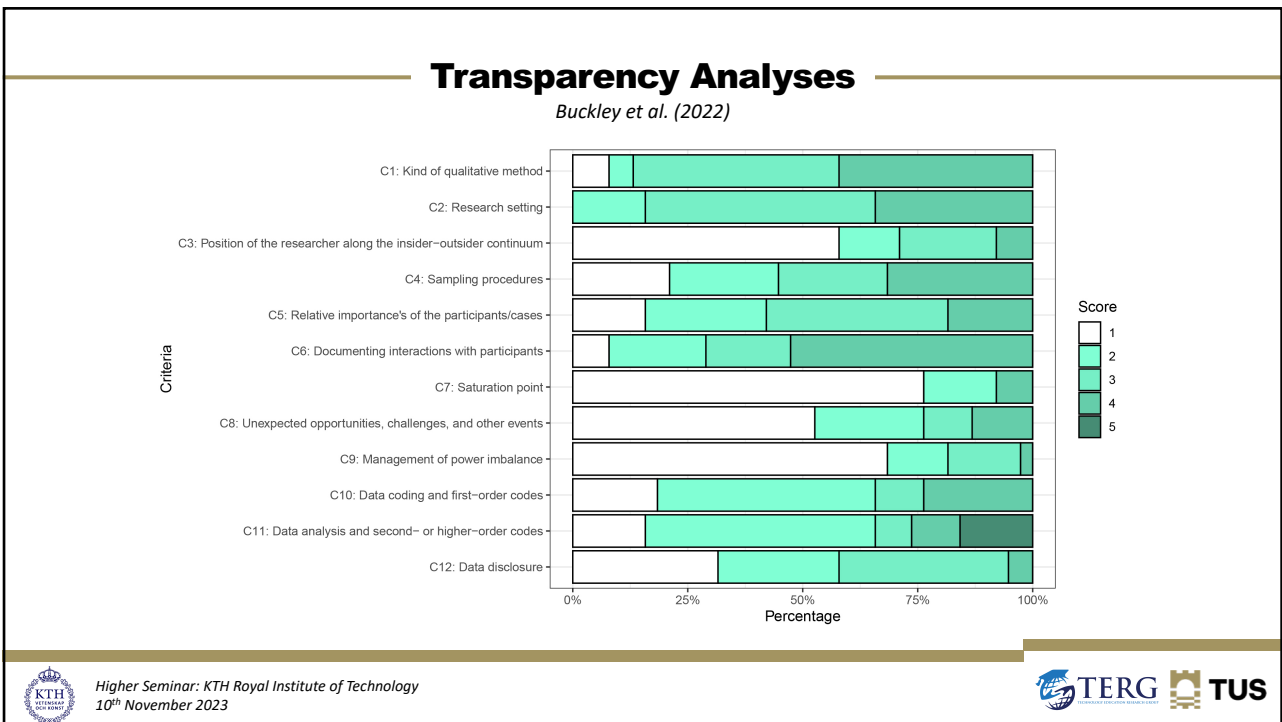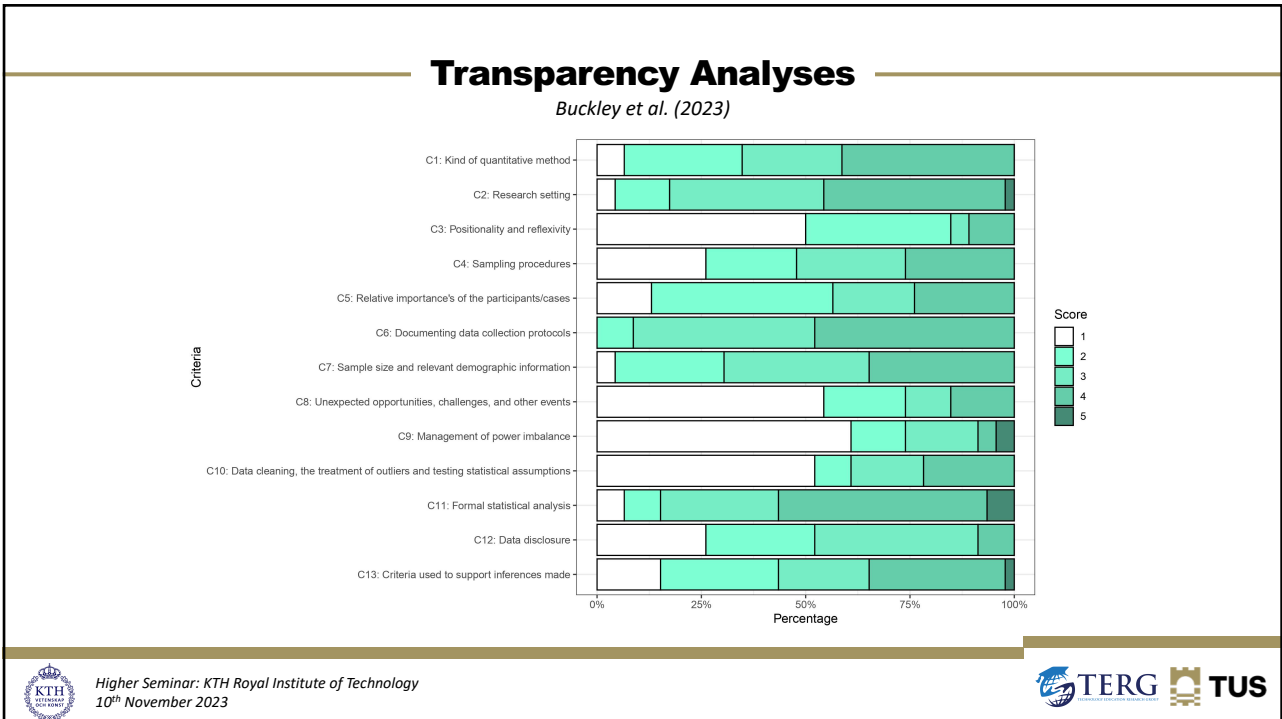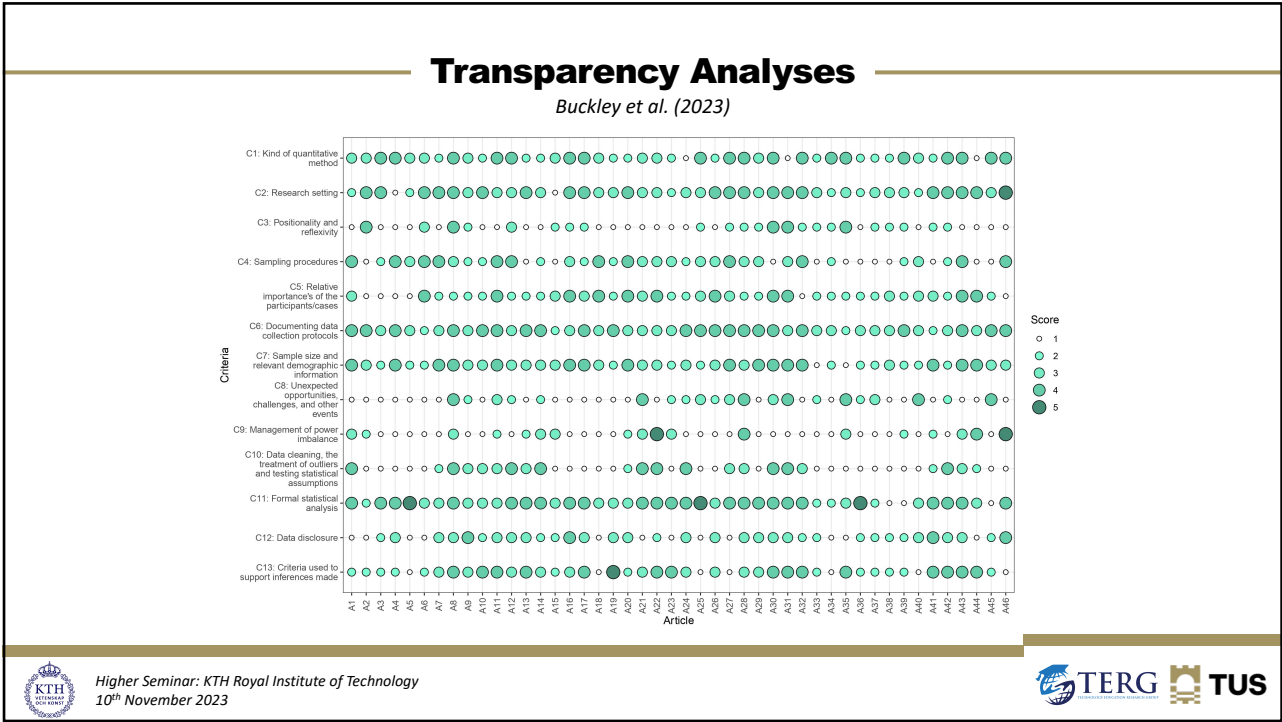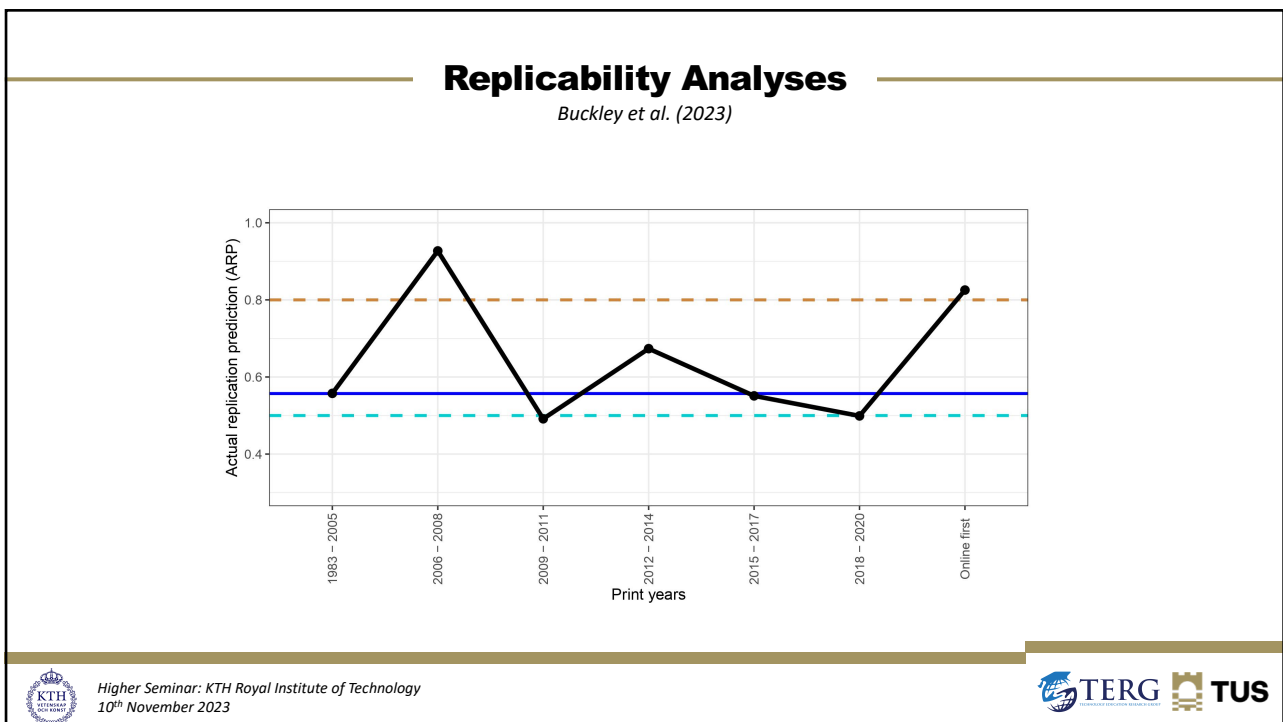*Higher Seminar: KTH Royal Institute of Technology*
10th November 2023

"Some" implications

Academics/Research Students → Literature Review → Extension Study / Replication Study

"Some" implications

Academics/Research Students → Literature Review → Translation to Practice

One Solution

A Second Solution: Research Credibility

LeBel et al. (2018)

Transparency	Reproducibility	Robustness	Replicability
Is the work clearly reported?	Could the data analytic procedure be reproduced exactly to come to the same substantive finding?	Is the primary substantive finding robust to different analytic procedures?	Would the primary substantive finding replicate in an independent replication study?

Technology Education Research Growth

Buckley et al. (2023)

AJTE = Australian Journal of Technology Education. DATE = Design and Technology Education: An International Journal.
IJTDE = International Journal of Technology and Design Education. JTE = Journal of Technology Education.
RSTE = Research in Science and Technological Education. Online first articles relate to articles published up to Tuesday 11th August 2020.

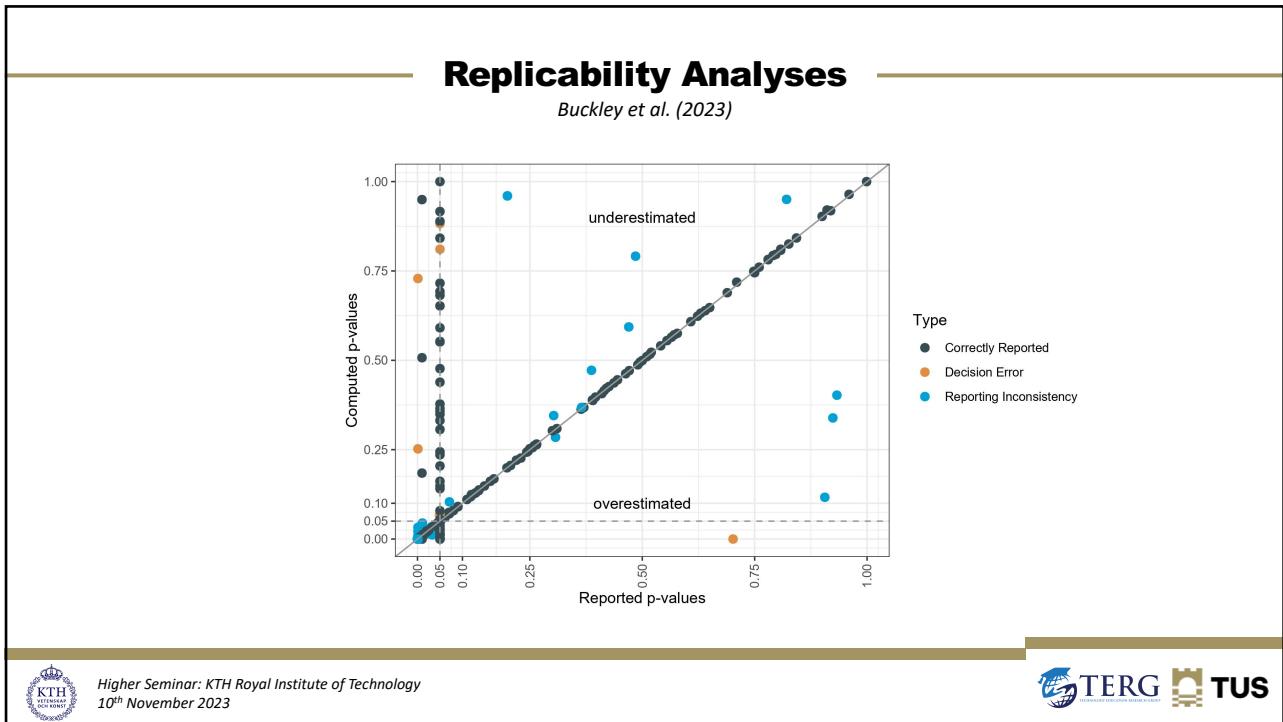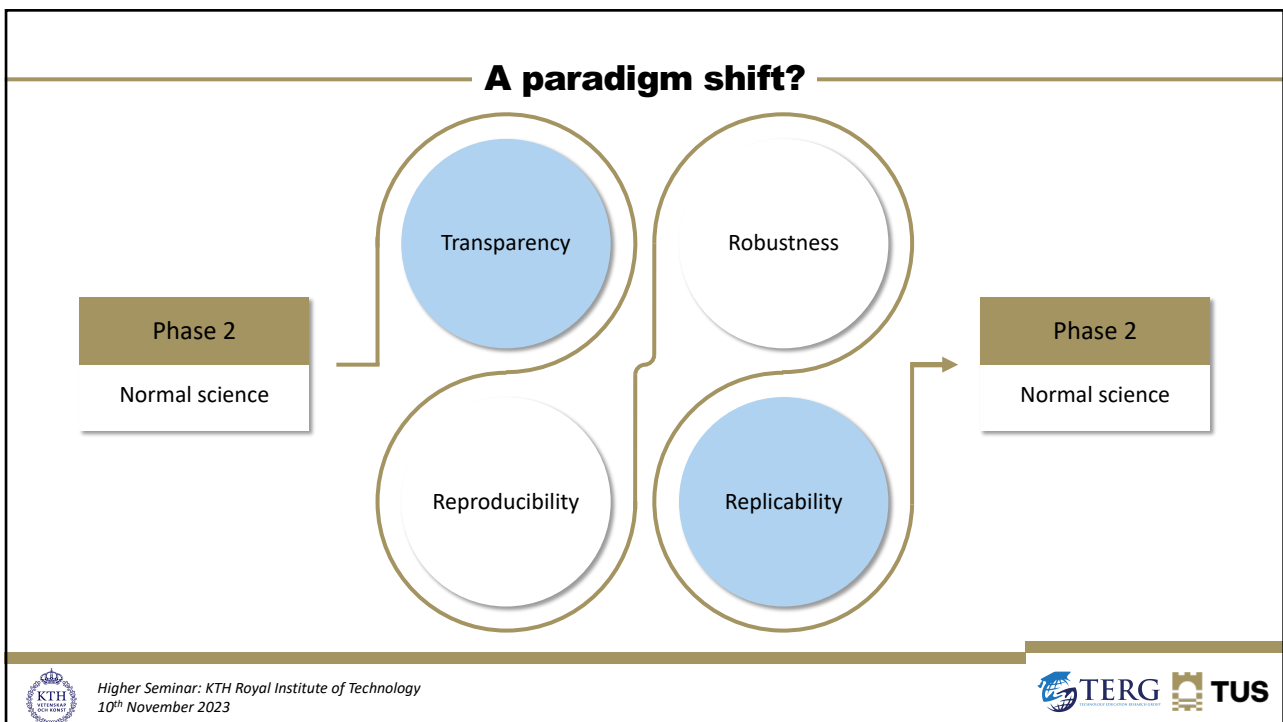*Higher Seminar: KTH Royal Institute of Technology*
10th November 2023

Technology Education Research Growth

Describing

Generalising

1980's &
1990's

2000's

2010's

2020's

Framing

Interrogating
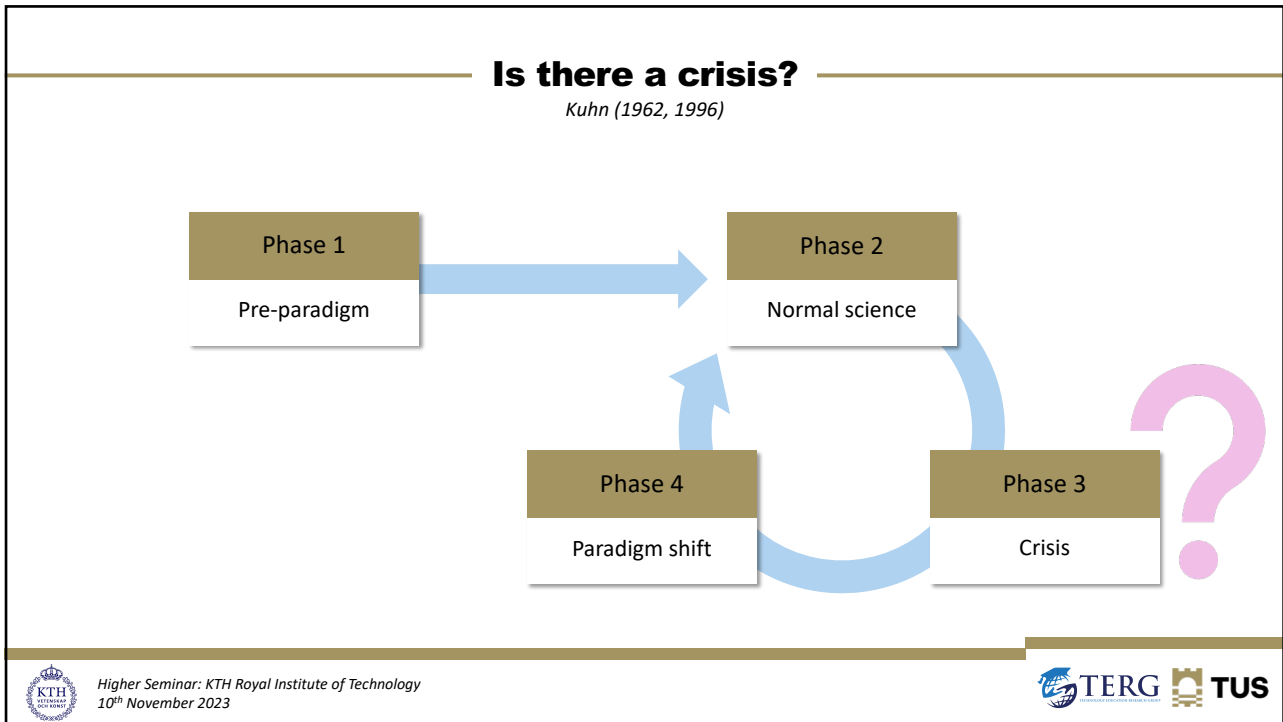
Higher Seminar: KTH Royal Institute of Technology
10th November 2023

Research Field Development
Kuhn (1962, 1996)

Phase 1
Pre-paradigm

Phase 2
Normal science

Phase 4
Paradigm shift

Phase 3
Crisis

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

TERG TUS

Research Credibility
LeBel et al. (2018)

Transparency	Reproducibility	Robustness	Replicability
Is the work clearly reported?	Could the data analytic procedure be reproduced exactly to come to the same substantive finding?	Is the primary substantive finding robust to different analytic procedures?	Would the primary substantive finding replicate in an independent replication study?

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

TERG TUS

Transparency Analyses
Buckley et al. (2022)

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

Transparency Analyses
Buckley et al. (2022)

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

Transparency Analyses
Buckley et al. (2023)

Transparency Analyses
Buckley et al. (2023)

Replicability Analyses
Buckley et al. (2023)

Replicability Analyses
Buckley et al. (2023)

Is there a crisis?

Kuhn (1962, 1996)

A paradigm shift?

Transparency criteria

Transparency criterion	Authors should (where possible and ethical)…
Kind of qualitative method	be explicit about what specific kind of qualitative method has been implemented (e.g., narrative research, grounded theory, ethnography, case study, phenomenological research)
Research setting	provide detailed information regarding contextual issues regarding the research setting (e.g., power structure, norms, heuristics, culture, economic conditions)
Position of the researcher along the insider-outsider continuum	provide detailed information regarding the researcher's position along the insideroutsider continuum (e.g., existence of a preexisting relationship with study participants, the development of close relationships during the course of data collection)
Sampling procedures	be explicit about the sampling procedures (e.g., theoretical sample, purposive sample, snowballing sample, stratified sample)
Relative importance of the participants/cases	be explicit about the contribution that key informants made to the study
Documenting interactions with participants	document interactions with participants (e.g., specify which types of interactions led to the development of a theme)
Saturation point	identify the theoretical saturation point and describe the judgment calls the researcher made in defining and measuring it
Unexpected opportunities, challenges, and other events	report what unexpected opportunities, challenges, and other events occurred during the study, how they were handled (e.g., participants dropped out of the study, a new theoretical framework was necessary), and implications
Management of power imbalance	report and describe whether power imbalance exists between the researcher and the participants and how it was addressed (e.g., endorsement from a prestigious institution, self-acquaintance, asking sensitive questions)
Data coding and first-order codes	be clear about the type of coding strategies adopted (e.g., structural, in vivo, open/initial, emotional, vs.)
Data analysis and second- or higher-order codes	how the data were analyzed (e.g., focused, axial, theoretical, elaborative, longitudinal)
Data disclosure	make raw materials available (e.g., transcripts, video recordings)

Higher Seminar: KTH Royal Institute of Technology
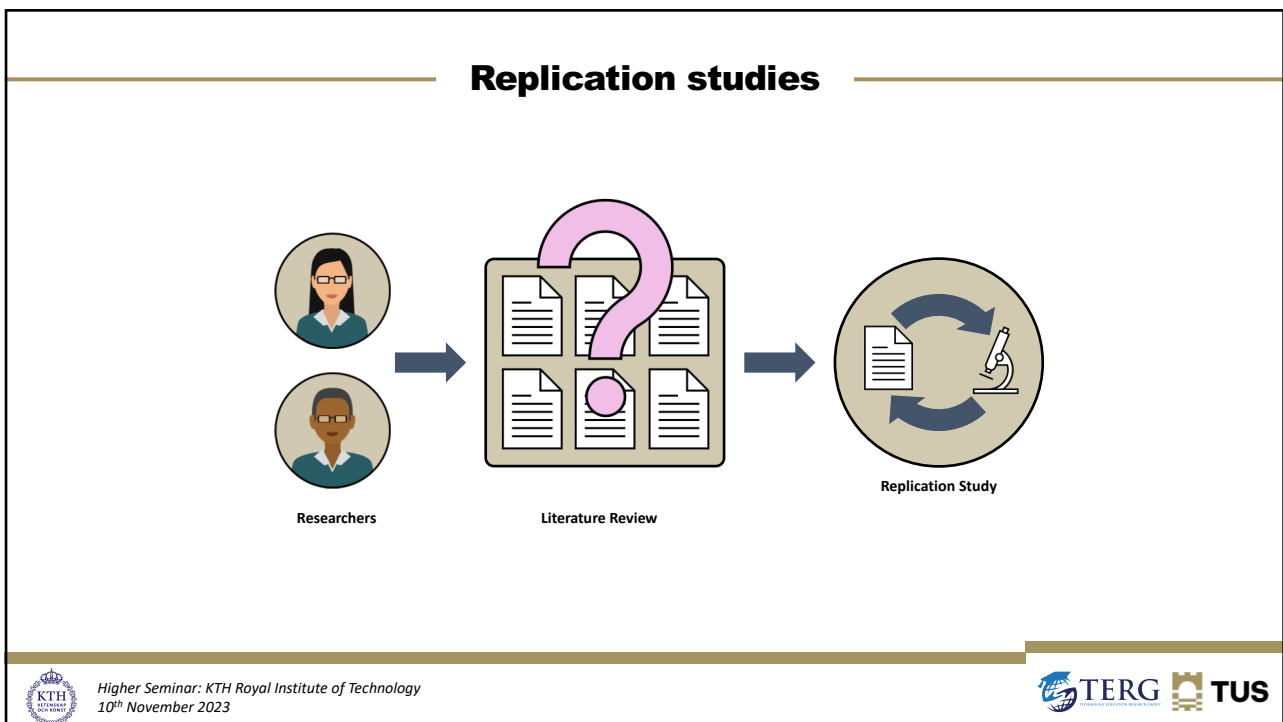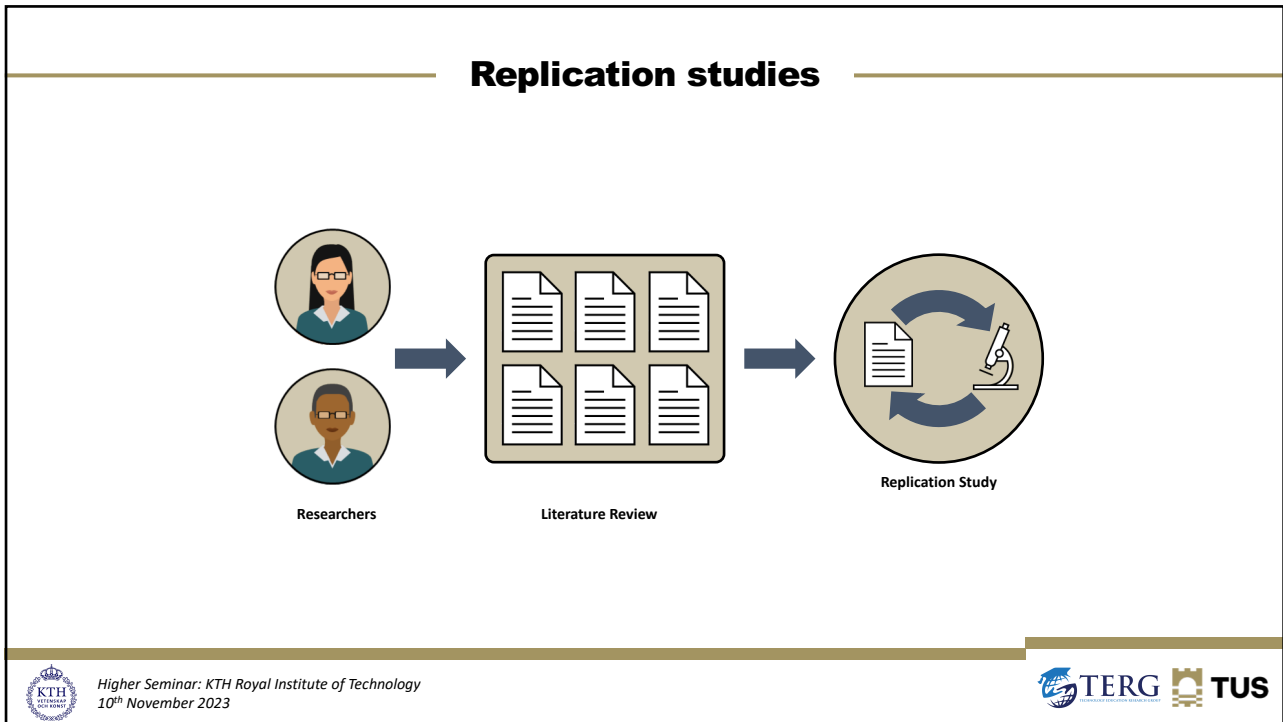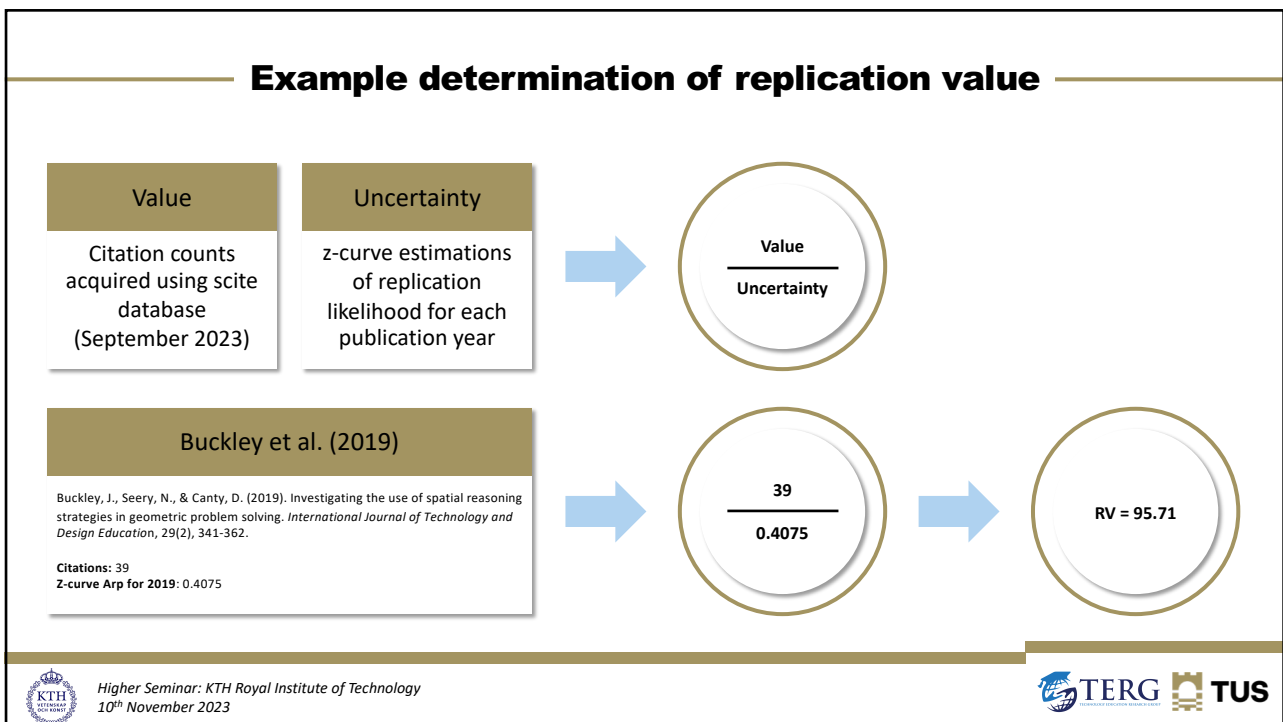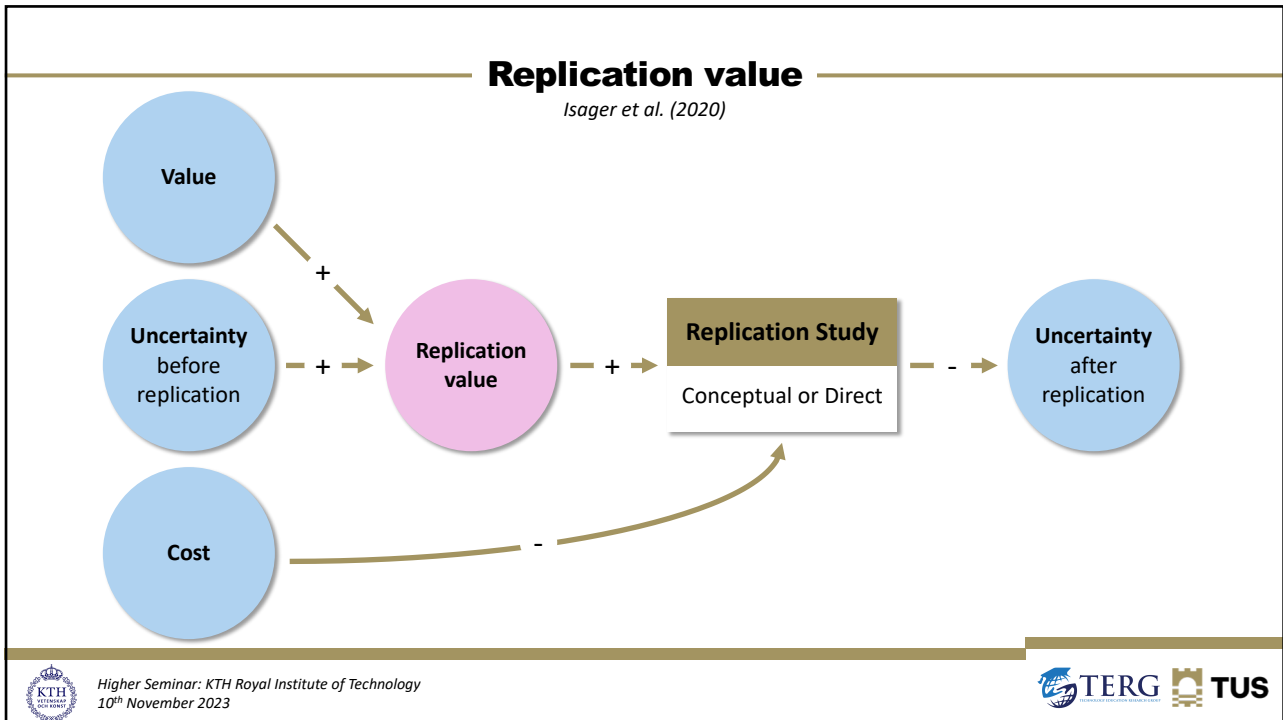*10th November 2023*

TERG TUS

Transparency criteria

| Transparency criterion	Authors should (where possible and ethical)…
Kind of quantitative method	be explicit about the particular quantitative methodology used in the study (e.g., an experimental design such as a true experiment or quasi-experiment, or a survey methodology such as longitudinal or cross-sectional design).
Research setting	provide detailed information about the physical, social, and cultural milieu of the study (e.g., the lab setting for a clinical trial, or location or demographic details of participants in an effectiveness study) and why these conditions were important to permit the identification or creation of similar settings in replication attempts.
Sampling procedures	be explicit about the procedures used to select participants or cases for the study (e.g., convenience, purposive, theoretical, random), and processes and tools used to assign participants to different conditions if applicable.
Relative importance of the participants/cases	be explicit about the study's sample and the relative importance of each participant/case or included demographic groups and representativeness.
Documenting data collection protocols	provide details and descriptions of methodological instruments used in the collection or gathering of data, to include details of instrument validation and reliability from previous studies.
Data reliability	provide detailed information regarding the reliability of all data to include computed statistics of reliability for objective instruments (e.g., psychometric tests) and computed statistics of interrater reliability where subjective scoring was undertaken (e.g., for educational assessments). This should include clear and replicable justification of coders.
Sample size and relevant demographic information	be explicit about the size and characteristics of the study sample, and how the sample size was determined.
Unexpected opportunities, challenges, and other events	provide detailed information about any unexpected opportunities, challenges, and events that occur during all stages of the research process.
Management of power imbalance, and incentivisation	report and describe whether a power imbalance exists between the researcher(s) and participants and how it was addressed. Also note any participation incentives which were used in the recruitment of participants or declare that none were used.
Data cleaning, the treatment of outliers and testing statistical assumptions	be explicit about the approach and steps taken in pre-processing or tidying collected data prior to formal analysis (e.g., data cleaning procedures or dealing with missing data, steps taken to identify and treat potential outliers, and approaches to testing statistical assumptions relating to statistical tests to be conducted).
Formal statistical analysis	provide detailed information about and use of formal statistical tests (e.g., univariate and multivariate tests) including describing the input data and test results.
Data disclosure	provide the raw material including any information collected by the researcher before any manipulation (i.e., analysis) (e.g., survey responses or test scores).
Accepted standard of evidence	be explicit about, and justify, the criteria which must be met for something to be judged as evidence supporting an inference (e.g., an alpha value of 0.05 against which p-values are compared, and adjustments for multiple comparisons)

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

TERG TUS

Replication studies

Researchers

Literature Review

Replication Study

Replication studies

Researchers

Literature Review

Replication Study

Replication value

Isager et al. (2020)

Value

Uncertainty before replication

Cost

Replication value

Replication Study — Conceptual or Direct

Uncertainty after replication

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

Example determination of replication value

Value — Citation counts acquired using scite database (September 2023)

Uncertainty — z-curve estimations of replication likelihood for each publication year

$$\frac{Value}{Uncertainty}$$

Buckley et al. (2019)

Buckley, J., Seery, N., & Canty, D. (2019). Investigating the use of spatial reasoning strategies in geometric problem solving. *International Journal of Technology and Design Education*, 29(2), 341-362.

Citations: 39
Z-curve Arp for 2019: 0.4075

$$\frac{39}{0.4075}$$

RV = 95.71

Higher Seminar: KTH Royal Institute of Technology
10th November 2023

How could we progress?
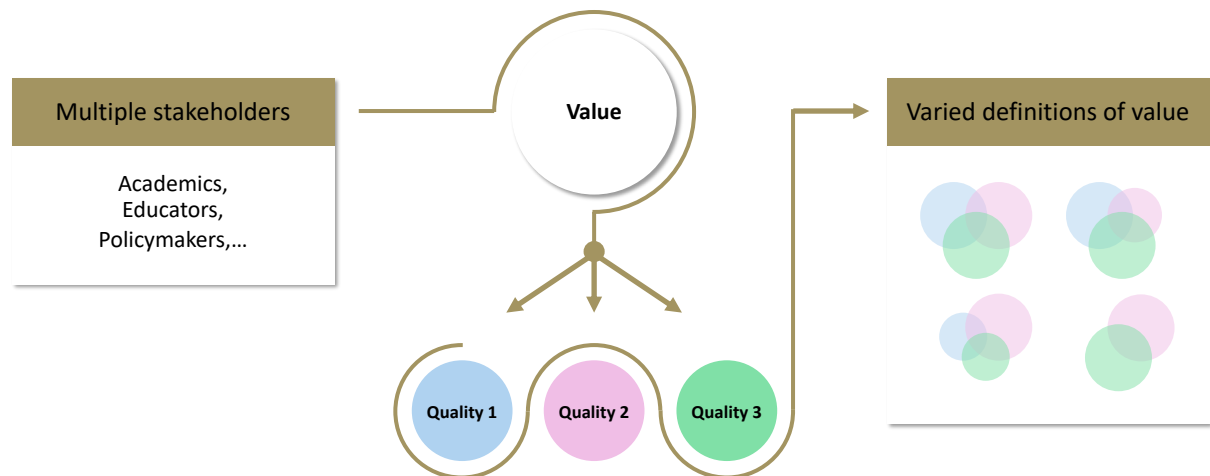
References

Buckley, J., Adams, L., Aribilola, I., Arshad, I., Azeem, M., Bracken, L., Breheny, C., Buckley, C., Chimello, I., Fagan, A., Fitzpatrick, D. P., Garza Herrera, D., Gomes, G. D., Grassick, S., Halligan, E., Hirway, A., Hyland, T., Imtiaz, M. B., Khan, M. B., … Zhang, L. (2022). An assessment of the transparency of contemporary technology education research employing interview-based methodologies. *International Journal of Technology and Design Education*, 32(4), 1963–1982.

Buckley, J., Araujo, J. A., Aribilola, I., Arshad, I., Azeem, M., Buckley, C., Fagan, A., Fitzpatrick, D. P., Garza Herrera, D. A., Hyland, T., Imtiaz, M. B., Khan, M. B., Lanzagorta Garcia, E., Moharana, B., Mohd Sufian, M. S. Z., Osterwald, K. M., Phelan, J., Platonova, A., Reid, C., … Zainol, I. (2023). How transparent are quantitative studies in contemporary technology education research? Instrument development and analysis. *International Journal of Technology and Design Education*. Buckley, J., Hyland, T., & Seery, N. (2023). Estimating the replicability of technology education research. *International Journal of Technology and Design Education*, 33(4), 1243–1264.

Buckley, J., Hyland, T., & Seery, N. (2023). Estimating the replicability of technology education research. *International Journal of Technology and Design Education*, 33(4), 1243–1264.

Isager, P. M., Aert, R. C. M. van, Bahník, Š., Brandt, M., DeSoto, K. A., Giner-Sorolla, R., Krueger, J., Perugini, M., Ropovik, I., Veer, A. van 't, Vranka, M. A., & Lakens, D. (2020). Deciding what to replicate: A formal definition of "replication value" and a decision model for replication study selection. In *MetaArXiv*.

Kuhn, T. (1996). *The structure of scientific revolutions*. Chicago University Press.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402.

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology*, 65(11), 2271–2279.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 943. https://doi.org/10.1126/science.aac4716

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530. Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., … Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356.

Thank you!

Jeff Buckley
Email: Jeffrey.buckley@tus.ie

TUS