Conferences
leibniz-psychology.org

Research Synthesis

June 10-12, 2018
Trier, Germany

Abstract Collection

leibniz-psychology.org

Pre-Conference Workshop

Howard White, Bernd Weiss

"Introduction to Systematic Reviews: a half-day workshop."

Sunday, June 10
9 am - 12 pm

This event is offered in collaboration with Campbell Collaboration and GESIS Leibniz Institute for the Social Sciences.

Background. The number of research publications doubles every nine years. More than 6,000 new journals are started every year. And policy makers do not read academic journals anyway. So how are we to stay on top of the current state of academic literature, what can we learn and how can consensus positions emerge from a vast array of differing studies in different contexts with different findings, and how can policy be informed by evidence?

The adoption of evidence-based medicine was driven by systematic reviews. Systematic reviews summarize all available high quality evidence addressing a specific question. Non-systematic reviews, including traditional literature reviews, are more prone subject to bias from missing studies and selective reporting. Well conducted reviews deliver clear policy messages. The evidence-based policy movement promotes the adoption of the systematic review approach as a standard methodology in social and economic analysis. This half-day hands on workshop is introduces participants to the principles of systematic reviews.

Workshop format. The workshop has an interactive format of lectures followed by hands-on sessions. There will be a short quiz at the end. The hands-on sessions are partly group work and partly individual exercises. The group work will be based on developing a research question using the Campbell Collaboration Title Registration Form. Each statistical lecture is followed by individual exercises, where participants perform meta-analytical analyses using R or Stata.

Who is the workshop for? The workshop is for early to mid-career researchers interested in undertaking a systematic review. Prior experience in conducting a review is not necessary. Participants should: (1) have a good grounding in statistics/econometrics, and be familiar with approaches to estimating effects/impact in the presence of selection bias; (2) be familiar with R or Stata. Participants should bring their own laptop to the workshop.

Session 1: **"The use of evidence synthesis to improve policy and practice: International perspectives"**

Howard White

Sunday, June 10
2:30 pm - 4 pm

Session 2: **Applications: I&O-Psychology**

Sunday, June 10
4:30 pm - 6 pm

4:30 pm - 5:15 pm	*Long Zhang* "How ethical leadership impacts employee organizational citizenship behavior? A meta-analytic review of competing mediating mechanisms based on two stage meta-analytic structural equations modeling (TSSEM)."
5:15 pm - 6 pm	*Juan David Reyes-Gómez, Pilar López-Belbeze, Josep Rialp* "The effects of strategic orientations on firm performance, and the mediating role of innovation outcomes: a meta-analytic path analysis."

4:30 pm - 5:15 pm

Long Zhang

"How ethical leadership impacts employee organizational citizenship behavior? A meta-analytic review of competing mediating mechanisms based on two stage meta-analytic structural equations modeling (TSSEM)."

Background. Extensive research has shown that ethical leadership plays a key role in shaping employee behaviors (Loi, Lam, & Chan, 2012; Neubert, Carlson, Kacmar, Roberts, & Chonko, 2009; Yong-jun, 2012). Although prior study examined the explanatory mechanism of ethical leadership on organizational citizenship behavior or OCB from the perspectives of cognitive and affective trust (Lu, 2014), the alternative perspectives, such as justice perspective, for explaining this relationship were missing. The core premise of justice perspective is the perception of fairness (Folger, 2001; Folger & Cropanzano, 2001). Employee's perception of fairness is captured by the construct of organizational justice, which contains three aspects: distributive justice, interactional justice and interpersonal justice. Distributive justice regarding fairness in outcome allocation, procedural justice regarding fairness in procedures applied to make allocations, and interactional justice regarding interpersonal treatment experienced by focal employees (Cohen-Charash & Spector, 2001; Colquitt, Conlon, Wesson, Porter, & Ng, 2001). However, how these three types of justice differ from each other, if any, in explaining the relationship between ethical leadership and OCB is even a more underexplored research question. To address this important research question and to advance the investigation of the mechanisms in ethical leadership-OCB linkage, we conducted a meta-analytic review using two stage meta-analytic structural equations modeling (TSSEM) to examine and

compare the mediating roles of different justice mechanisms.

Objectives. An interesting extension from both theoretical and practical standpoints is that among three dimensions of justice, whether one lens is better than the others in explaining how ethical leadership renders employee behavioral reactions in terms of the changes in OCBI and OCBO. Comparing different justice mechanisms, including interactional justice, distributive justice and procedural justice, enables us to better understand the unique processes underlying distinct justice aspects, and to facilitate the advancement of ethical leadership research.

Research question and hypotheses. While existing research have explained the effects of ethical leadership from various perspectives, such as cognitive and affective trust (Lu, 2014), and organizational politics (Kacmar, Andrews, Harris, & Tepper, 2013), there is a critical omission in exploring its mediational pathways, that is, the perspective of justice. Further, it is also important to empirically compare the explanatory powers of the three distinct dimensions of justice and to identify which pathway can better explain the influence of ethical leadership on employee OCBI and OCBO (Shin, 2012). Therefore, this research aims to address the following research question and three hypotheses:

Research Question: From the perspective of justice, which mechanism, among interactional justice, distributive justice or procedural justice, has stronger exploratory power regarding how ethical leadership influences employee OCBI and OCBO?

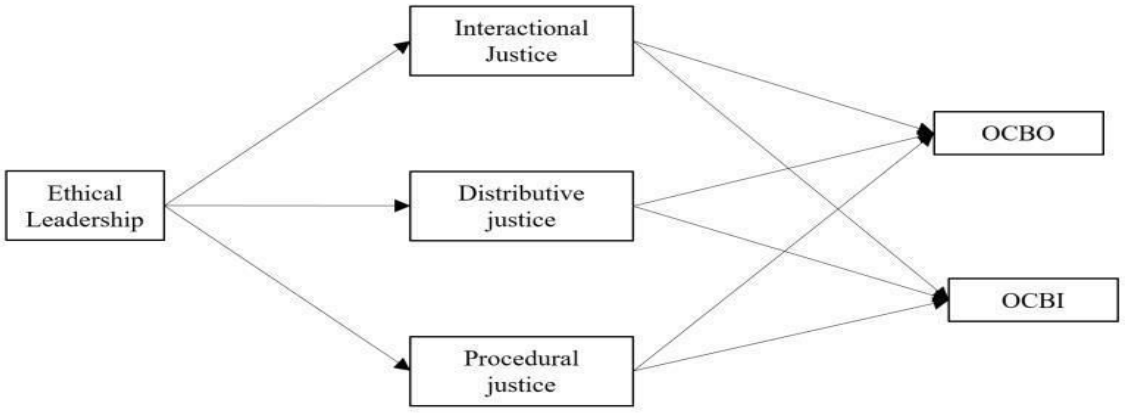
Hypothesis 1: Ethical leadership is positively related to employee organizational justice, including interactional justice, distributive justice and procedural justice.

Hypothesis 2: Organizational justice, including interactional justice, distributive justice and procedural justice, mediates the relationship between ethical leadership and employee OCB (OCBO, OCBI) respectively.

Hypothesis 3: Distributive justice and procedural justice have stronger explanatory power regarding how ethical leadership influence OCBO; interactional justice is more exploratory regarding why ethical leadership influence OCBI.

The theoretical model is presented as below:

Figure 1: Proposed model



Method/Approach. We applied two stage meta-analytic structural equations modeling (TSSEM) to test our hypothesized model and research question, which is more rigorous than traditional meta-analytic structural equations modeling (MASEM) (Viswesvaran & Ones, 1995). Based on the standard procedure of TSSEM proposed by Cheung (2015), a correlation matrix is pooled in the first stage, which is used to fit structural equation models in the second stage. The data homogeneity is tested. Fixed-effects model will be applied if the population effect sizes are homogeneous. Alternatively, the random-effects model will be applied if the homogeneity of effect sizes is not satisfied. In the second stage, a vector of the pooled correlation matrix ρˆand its asymptotic sampling covariance matrix Vˆ are estimated, which is used to fit the SEM. The R package metaSEM is applied to analyze the data. TSSEM method offers several advantages. First, MASEM enables us to test our theoretical model and get more accurate estimates by simultaneously examining multiple mediators and outcomes (Clarke, 2005; Viswesvaran & Ones, 1995). Third, TSSEM helps us to compare the explanatory power of two different pathways (i.e., organizational justice and work stress) by comparing alternative models (Cheung & Chan, 2005; Mackey, Frieder, Brees, & Martinko, 2015; Rosenthal & Dimatteo, 2001). Third, TSSEM can also be used to solve the redundant model problem and to get consistent results by comparing different possible models in a systematic manner and thereby providing evidence of the most fitted model (Rosenthal & Dimatteo, 2001).

Results/Findings. Using TSSEM, the results based on 234 studies, comprising 84,505 participants, indicated that organizational justice mediated the link between ethical leadership and employee OCB. The results further revealed that distributive justice and procedural justice perspective accounted more for the effect of ethical leadership on employee OCBO, while the interactional justice perspective accounted more for the impact of ethical leadership on employee OCBI.

Conclusions and Implications (expected). Drawing upon justice theoretical perspectives,

our meta-analytic review examined and explained how ethical leadership influenced employee OCB. Further, we uncovered which theoretical lens, among interactional justice, distributive justice and procedural justice, accounted more for the respective relationships.

References

Cheung, M. 2015. Meta-analysis: A structural equation modeling approach: John Wiley & Sons.

Cheung, M. W. L., & Chan, W. 2005. Meta-analytic structural equation modeling: A two-stage approach. Psychological Methods, 10(1): 40-64.

Clarke, K. A. 2005. The phantom menace: Omitted variable bias in econometric research. Conflict Management and Peace Science, 22(4): 341-352.

Cohen-Charash, Y., & Spector, P. E. 2001. The role of justice in organizations: A meta-analysis. Organizational Behavior and Human Decision Processes, 86(2): 278-321.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. 2001. Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology, 86: 425-445.

Folger, R. 2001. Fairness as deonance. In S. Gilliland, D. Steiner, & D. Skarlicki (Eds.), Theoretical and cultural perspectives on organizational justice: 3-33. Greenwich, CT: Information Age.

Folger, R., & Cropanzano, R. 2001. Fairness theory: Justice as accountability, Advances in organization justice.: 1-55: Stanford University Press.

Kacmar, K. M., Andrews, M. C., Harris, K. J., & Tepper, B. J. 2013. Ethical leadership and subordinate outcomes: The mediating role of organizational politics and the moderating role of political skill. Journal of Business Ethics, 115(1): 33-44.

Loi, R., Lam, L. W., & Chan, K. W. 2012. Coping with job insecurity: The role of procedural justice, ethical leadership and power distance orientation. Journal of Business Ethics, 108(3): 361-372.

Lu, X. 2014. Ethical leadership and organizational citizenship behavior: The mediating roles of cognitive and affective trust. Social Behavior and Personality: an international journal, 42(3): 379-389.

Mackey, J. D., Frieder, R. E., Brees, J. R., & Martinko, M. J. 2015. Abusive Supervision:

A Meta-Analysis and Empirical Review. Journal of Management. 43 (6): 1940-1965.

Neubert, M. J., Carlson, D. S., Kacmar, K. M., Roberts, J. A., & Chonko, L. B. 2009. The virtuous influence of ethical leadership behavior: Evidence from the field. Journal of Business Ethics, 90(2): 157-170.

Rosenthal, R., & Dimatteo, M. R. 2001. Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. Annual Review of Psychology, 52(1): 59-82.

Shin, Y. 2012. CEO ethical leadership, ethical climate, climate strength, and collective organizational citizenship behavior. . Journal of Business Ethics, 108(3): 299-312.

Viswesvaran, C., & Ones, D. S. 1995. Theory testing: Combining psychometric meta-analysis and structural equations modeling. Personnel Psychology, 48(4): 865-885.

Yong-jun, Z. 2012. The Influence of Ethical Leadership on Employees' CWB: from Social Learning and Social Exchange Perspective. Journal of Business Economics, 12: 005.

5:15 pm - 6 pm

Juan David Reyes-Gómez, Pilar López-Belbeze, Josep Rialp

"The effects of strategic orientations on firm performance, and the mediating role of innovation outcomes: a meta-analytic path analysis."

Under the resource-based view of the firm (RBV) the strategic orientation of a firm is an organizational resource and capability embedded into organizational culture which refers to the means by which firms choose to attempt to create a sustainable presence in the markets in which they compete. This concept has attracted widespread attention in the marketing, management and entrepreneurship literature over the past two decades, focusing mainly in three orientations: market orientation (MO), entrepreneurial orientation (EO), and learning orientation (LO).

A large number of studies on strategic orientations concentrated on their individual effects on firm performance, finding MO, EO and LO to directly influence firm performance independently and parallelly, whereas a fewer research body jointly and in an interrelated perspective studied MO, EO, and LO direct effects on performance. Research attention also focused on mediating mechanisms as innovation between the strategic orientation and firm performance relationship.

It is plausible that the impact of strategic orientations has been understated due to their effects may be indirect through innovation or, which is the same, innovation could mediate the relation. Three different hypothesized approaches –depicted as theoretical models–

assumed in past research are identified:

1) the universalistic approach which implies that the strategic orientations exert independent and parallel direct effects on firm performance;
2) the synergistic approach which implies that strategic orientations complementarily and jointly exert indirect effects on firm performance through innovation as full mediator in the relationship;
3) the holistic approach which implies that strategic orientations complementarily and jointly exert simultaneous direct and indirect effect on firm performance through innovation as partial mediator.

This study aims to quantitatively synthetize the available literature's data in an integrative meta-analytic path analysis framework allowing for theory testing, to assess the usefulness and validity of the hypothesized approaches using the unified MASEM's two-stage structural equation modeling. In this sense, two research questions are addressed: which competing model extracted from the literature on strategic orientations, innovation and firm performance fits better the meta-analytic data? Does innovation play a mediating role –whether null, full or partial– in the strategic orientations and firm performance relationship?

The dataset consisted of 119 independent samples extracted from 116 selected empirical articles yielding a total sample of 32.322 observations.

Findings indicate that the holistic hypothesized approach fits better the cumulated data. Thus, the universalistic approach may possible be abandoned and replaced with the idea that the fit between strategic orientations and firm performance is dependent on the interplay between strategic orientations, and innovation plays a mediating role, considered in the synergetic and holistic approaches. The latter approach is more effective than the former one in a theoretical modeling perspective due to its suitability for linking more complex relationships. Within the holistic approach it is demonstrated that innovation outcomes partially mediate the relationship between EO and firm performance, whereas fully mediates the MO and LO relationships with firm performance. Managers should be aware that achieving superior performance in competitive markets mainly depends on the pursuing of market opportunities through the delivery of successful innovations given a proper combination of orientations and the synergies generated by 1) identifying and exploiting new products or markets, 2) the focus on creating value for customers anticipating their latent needs, and 3) the continuous learning that provides the vision to predict what the market may become. Future meta-analytic path analysis may include implementation of study characteristics as moderators, such as size of the firm (large vs. SME), industry type (manufacturing vs. services) and regional culture (individualism vs. collectivism). Using random-effects subgroup analysis is strongly suggested.

Session 3: **Methodological Issues (Part 1)**

Monday, June 11
10:30 am - 12 pm

10:30 am - 10:50 am	*Loukia Maria Spineli, Juan Jose Yepes-Nuñez, Holger Schünemann* "A systematic survey shows that reporting and handling of missing outcome data in networks of interventions is poor."
10:50 am - 11:10 am	*Lukasz Stasielowicz, Timo Gnambs* "Correcting for measurement error using meta-regression analyses: A comparison of approaches."
11:20 am - 11:40 am	*Melissa Bond, Katja Buntins, Svenja Bendelier, Michael Kerres, Olaf Zawacki-Richter* "Fuzzy concepts and large literature corpi: Addressing methodological challenges in systematic reviews."
11:40 am - 12 pm	*Martin Voracek, Michael Kossmeier, Ulrich S. Tran* "A specification-curve and multiverse-analysis approach to meta-analysis."

10:30 am - 10:50 am

Loukia Maria Spineli, Juan Jose Yepes-Nuñez, Holger Schünemann

"A systematic survey shows that reporting and handling of missing outcome data in networks of interventions is poor."

Background. Relevant review studies on Cochrane and conventional systematic reviews with meta-analyses have already revealed serious inadequacies in the reporting and handling of missing outcome data (MOD). To our knowledge, there is no such relevant survey specific to systematic reviews with multiple interventions.

Objectives. To provide empirical evidence about prevalence, reporting and handling of MOD in systematic reviews with network meta-analysis (NMA) and acknowledgement of their impact on the inferences.

Research question(s) and/or hypothesis/es. (i) What is the reporting quality on MOD in systematic reviews with NMA; (ii) whether and how the authors of systematic reviews with multiple interventions handle MOD; (iii) whether the authors of systematic reviews with multiple interventions acknowledge the implications of MOD on the interpretation and discussion of the results.

Method/Approach. We conducted a systematic survey including all published systematic reviews of randomized controlled trials and at least three interventions from January 1, 2009 until March 31, 2017.

Results/Findings. We retrieved 387 systematic reviews with NMA. Description of MOD

was available in 63 reviews. Intention-to-treat analysis was the most prevalent method (71%), followed by MOD investigated as secondary outcome (e.g. acceptability) (40%). Bias due to MOD was evaluated in half the reviews with explicit judgments in 18 (10%) reviews. Only 88 reviews interpreted their results acknowledging the implications of MOD and mostly using the NMA results on MOD as secondary outcome. We were unable to judge the actual strategy applied to deal with MOD in 65% of the reviews due to insufficient information. Six percent of NMAs were re-analysed in sensitivity analysis considering MOD, while 4% explicitly justified the strategy for dealing with MOD.

Conclusions and implications. The description and handling of MOD as well as the acknowledgement of their implications in the inferences of NMA are deemed underreported. The conduct of systematic reviews with NMA will benefit from the development of comprehensive guidelines and education regarding the ubiquity and implications of missing outcome data and their impact on the quality of the conclusions delivered to the healthcare system.

10:50 am - 11:10 am

Lukasz Stasielowicz, Timo Gnambs

"Correcting for measurement error using meta-regression analyses: A comparison of approaches."

Background. Artifact corrections based on the Schmidt and Hunter approach (2015) allow meta-analysts to account for measurement error in study outcomes. This correction follows the well-known correction for attenuation in classical test theory. Typically, these corrections for unreliability of the instruments utilized in primary studies lead to larger effect sizes. Corrected effect sizes are often regarded as more precise estimates of the true relationship between variables than an uncorrected effect size. However, the assumption that measurement error leads to an underestimation of the effect size can be criticized, because an overestimation is also plausible. In other words, measurement error may affect both the magnitude and the sign of the estimate (Gelman & Carlin, 2014). Hox and colleagues (2017) suggested an alternative approach to account for measurement error. Specifically, the instrument reliability (e.g., coefficient alpha) can be included as a moderator variable in a meta-regression model. Accordingly, the pooled effect size corrected for measurement error can be inferred from the meta-regression results. More importantly, the corrected mean effect size can theoretically also become of smaller magnitude than the uncorrected estimate. Furthermore, the sign of the corrected mean effect size might also change. It is not possible to observe such patterns within the Hunter-Schmidt framework when correcting for unreliability.
Objectives. As the approach proposed by Hox and colleagues (2017) has not been

embraced to the same extent as the Hunter-Schmidt approach, it is not clear when the results of the two approaches diverge. Thus, some meta-analytic findings may be less robust than readers assume. The main objective of the present simulation study was to systematically compare both approaches under several scenarios.
Research questions. We hypothesized that the Hunter-Schmidt approach and the meta-regression approach will perform differently with respect to the bias of the meta-analytic estimate. Considering that effect sizes corrected for measurement error within the Hunter-Schmidt framework are of the same magnitude or larger than the uncorrected effect sizes we expected that the meta-analytic estimates based on the Hunter-Schmidt approach will be systematically larger than the true relationship.
Method. In order to compare the two approaches (Hunter-Schmidt, meta-regression) we juxtaposed their effect estimates with the true effect size in a simulation study. The simulated scenarios differed with respect to the number of primary studies, magnitude of the relationship, and instruments' reliability range.
Results and Conclusions. The Hunter-Schmidt approach yielded mean effect sizes that were somewhat larger than the true relationship, particularly in missing not at random scenarios. In general, the meta-regression approach of Hox and colleagues yielded more conservative estimates. We conclude that the two examined approaches to correction for measurement error can yield different estimates of bivariate relationships. Further research should include boundary conditions for each meta-analytic approach (e.g., if the effect size is related to the reliability). Furthermore, one could examine the performance of the two approaches when missing reliability values are imputed. In future studies, one could also examine other approaches, i.e. Bayesian meta-analysis with correction for measurement error.

11:20 am - 11:40 am

Melissa Bond, Katja Buntins, Svenja Bendelier, Michael Kerres, Olaf Zawacki-Richter

"Fuzzy concepts and large literature corpi: Addressing methodological challenges in systematic reviews."

Background

There are a lot of systematic reviews that analyze the impact of a specific educational technology tool or didactic approach on learning outcomes or student engagement, within a variety of educational settings (for an extensive discussion of systematic reviews as a method, see Gough et al., 2012). For example, Merchant et al. (2014) explored in a meta analysis the influence of virtual reality-based instruction within different learning

environments, Gao, Luo and Zhang (2012) summarized the research on microblogging in education, and Tess (2013) reviewed recent research of using social media in higher education.

All these questions are interesting and quite important for a scientific discussion. However, they are nevertheless not useful for the development of a suitable learning setting by practitioners, especially in light of the meta analysis by Tamim et al. (2011), who found that the use of learning technologies does not necessarily make learning more successful. This is not about the question of educational technology itself, but about the fit of learning objectives, pedagogy, learners and technology. The goal of our systematic review is therefore to identify the conditions under which the use of educational technology promotes student engagement in higher education.

Objectives

This very broad research question leads to several methodological challenges. First, student engagement is a multidimensional construct with a multitude of different facets (e.g. Kahu & Nelson, 2017). The specific aspects of student engagement that the various studies are particularly focused on are not explicitly discussed in a large portion of papers, which is relevant for the review, as they are embedded in a different theoretical context. As heterogeneous as the theoretical approaches are, so different are the research questions, the operationalization and the data analysis. Using our systematic review as an example, in our contribution we discuss the methodological challenges accounting to fuzzy concepts and large literature corpi in systematic reviews, and critically reflect on the approaches we used to address these challenges.

Research questions

Arising from our systematic review, two unusual methodological challenges have been encountered so far:

1. How to plan a systematic review when the constructs of interest are fuzzy?
2. What options are there to deal with a large article population?

Solutions/Approaches

Addressing fuzzy constructs

Contrary to previous studies that searched specifically for the term ‘engagement’ (e.g. Henrie, 2016), we decided to pursue a global search strategy. By creating the search strategy we define in a first step only the constructs ‘student’, ‘higher education’ and ‘technology’, but did not include the concept of student engagement itself. Through this, we wanted to make sure that we did not overlook relevant studies and account hereby for the fuzziness of the construct. Instead, the facets of student engagement and their

synonyms were identified by analyzing key theoretic papers in this area (Kahu, 2013; Filsecker & Kerres, 2014; Appleton, Christenson, & Furlong, 2008; Fredricks, Blumenfeld, & Paris, 2004; Mahatmya, Lohman, Matjasko, & Feldman Farb, 2012; Reeve, 2012; Skinner & Pitzer, 2012). We extracted from our pool of articles, the studies that contained the stem of one of the facets or their synonyms, and limited the period to be examined to articles from 2007. This reduced the initial article pool from 77,508 to 18,068. Four reviewers then screened the corpus titles and abstracts, excluding 13,916 due to predefined exclusion criteria (e.g. off topic, not in English, not an article, not empirical or not primary research).As the inclusion criteria in the area of student engagement are fuzzy, we compared our abstract screening results with three raters, until we reached an interrater reliability of . All abstracts where disagreement about exclusion remained, were discussed and resolved together. Of the 18,068 articles, a population of 4,152 articles to screen on full text remained at the end of initial screening.

Addressing a large literature corpus

Normally a full survey of all literature is carried out in systematic reviews. However, given the nature of limited time in project-based and funded work, and the demand to first screen the full texts of 4,153 articles and comprehensively code the ones for further inclusion, we needed to modify our approach. For this reason, we have used methods of sample size estimation in the social sciences (Kupper & Hafner, 1989). The goal here is to draw a sample that estimates the population parameters with a predetermined error range. For sampling, we used the R Package MBESS (Kelley, Lai, Lai & Suggets, 2018). When we accept a 5% error range, a percentage of an half and an alpha of 5%, we sampled 351 articles. We stratified the sample by the publishing year, because student engagement has become much more prevalent and educational technology has become more differentiated over the last ten years (Zawacki-Richter & Latchem, in review). After reviewing the 702 full texts considered initially, 124 articles were excluded, mostly because they did not focus on student engagement. We sampled the missing articles again from the stratified remaining population.

Discussion

Drawing the sample as part of a systematic review sets new ground in this field and provides a blueprint for synthesizing very heterogeneous and complex content in the field of Educational Technology. The sampling procedure, the resulting implications and the limitations are discussed in detail in the presentation. The fuzziness of central concepts of the review question and subsequently the heterogeneous body of studies are discussed as well in their relation to using the sampling as an innovative approach to reconcile large literature corpi whilst working systematically under time constraints.

11:40 am - 12 pm

Martin Voracek, Michael Kossmeier, Ulrich S. Tran

“A specification-curve and multiverse-analysis approach to meta-analysis.”

Session 4: **Applications: Clinical / Health Psychology**

Monday, June 11
2:30 pm - 4:30 pm

2:30 pm - 3 pm	*Karina Karolina Kedzior, Hannah Seehoff* "Common problems with meta-analysis in published reviews on major depressive disorders (MDD): a systematic review."
3 pm - 3:30 pm	*Bianca Annabelle Simonsmeier* "The positive effects of patient education on health: Insights from a second-order meta-analysis."
3:30 pm - 4 pm	*Dr. Katja Matthias, PD Dr. Matthias Perleth* "Science meets Reality - systematic reviews for health policy decisions according to new psychotherapeutic methods in Germany."
4 pm - 4:30 pm	*Bianca Annabelle Simonsmeier* "The relationship between subjective well-being and physical activity: A meta-analysis."

2:30 pm - 3 pm
Karina Karolina Kedzior, Hannah Seehoff

"Common problems with meta-analysis in published reviews on major depressive disorders (MDD): a systematic review."

Background

Meta-analysis is a useful tool for synthesising of empirical evidence necessary for decision and policy making in the clinical practice. Although enormously popular in the clinical field, there is little agreement as to how it should be conducted and how to assess its quality (Borenstein et al., 2009).

Objective. The objective of the current study is to explore the practice of meta-analysis in one clinically-relevant field (assessment of the clinical outcomes of various treatments on major depressive disorders, MDD) using a systematic review. We focus on the open-source journals published on the Biomedical Central (BMC) platform from Springer as sources of meta-analyses. BMC is chosen because the simple (and free) online access to its journals contributes to high readership rates (measured using an Altmetric Attention Score for each article), the scientific quality of its articles is assessed using the peer-review process, and the academic citations to its journals are traced using the impact factor scores. Following a preliminary search of the BMC website, three journals (BMC Psychology, BMC Psychiatry, Annals of General Psychiatry) were identified as relevant sources of meta-analyses in the field of MDD.

Research question. The research question in our study is: What is the practice of meta-analysis in reviews assessing the clinical outcomes of various treatments on MDD published in three BMC journals?

Method

The search, review selection, and data coding were done by each author independently and any inconsistencies were resolved by consensus.

Inclusion and exclusion criteria. The inclusion criteria for the current study were: 1) meta-analysis conducted using any method, 2) samples with MDD or any mood disorders, 3) clinical outcomes assessed immediately after treatment (acutely). The exclusion criteria for the current study were: 1) meta-analysis not conducted, 2) other diagnoses than MDD or healthy samples, 3) non-human data.

Search strategy. The electronic literature search of PubMed identified k=114 studies up to 14.06.2017 (Title: 'meta-analysis' AND Journal (BMC Psychiatry OR BMC Psychology OR Annals of General Psychiatry)). Inclusion criteria were met by k=14 studies (systematic reviews with meta-analysis) that were included in the final synthesis.

Coding procedures. The data in k=14 reviews were coded using two instruments: 1) the Assessment of Multiple Systematic Reviews scale (AMSTAR) focusing on the procedures and the quality of a systematic review (Shea et al., 2007), 2) the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist focusing on the reporting of methods and results in meta-analysis (Moher et al., 2009).

AMSTAR is a standardised 11-item scale with acceptable psychometric properties (Shea et al., 2007). The total AMSTAR score varies between 0 (minimum) to 11 (maximum quality of a systematic review). We used AMSTAR to assess the search strategy, study selection, data coding, and the risk of bias assessment in the k=14 reviews.

PRISMA is a standardised 27-item checklist on reporting of procedures and outcomes in a meta-analytic study (Moher et al., 2009). We supplemented Item 9 from AMSTAR ('Were the methods used to combine the findings of studies appropriate?') (Shea et al., 2007) with PRISMA items regarding Methods and Results (Moher et al., 2009) to assess the statistical procedures in a meta-analysis, including the overall analysis (model type, software, effect size type and computation, heterogeneity assessment), the sensitivity analysis (subgroup analysis, meta-regression, outlier analysis), and the publication bias analysis.

Results

Review characteristics. The k=14 studies (published in 2011-2017) were systematic reviews with meta-analysis regarding the acute clinical outcomes of various treatments on MDD symptoms or mood.

Systematic review quality. The k=14 reviews had a mean (±SD) AMSTAR score of 7±2 (range: 4-11). The majority of reviews received the AMSTAR scores for a duplicate study selection and data extraction, a comprehensive literature search, a list and characteristics of included primary studies, an assessment and a critical evaluation of study quality, and a statement regarding any conflict of interest. The majority of reviews lost the AMSTAR

scores due to the lack of or inadequate information about a priori protocol, consideration of unpublished sources, meta-analytical methods, or publication bias assessment.

Approach to meta-analysis. Meta-analysis was most frequently conducted as a random-effects model with inverse-variance weights in the RevMan software. However, 71% of meta-analyses were conducted inappropriately. Only a minority of meta-analyses used the correct effect sizes: standardised mean difference scores (pre-post) for scaled data (depression severity; 27% of studies) or odds ratios for nominal data (response or remission rates; 33%). Forest plots and heterogeneity statistics were missing from 21% of meta-analyses. Although various sensitivity analyses were conducted, including meta-regressions and subgroup analyses, most reviews reported univariate approaches using a very low volume of data (data from <10 studies). Publication bias was assessed in only 50% of reviews and mostly by visually inspecting funnel plots.

Conclusions and implications

Our study shows that 14 systematic reviews published in three BMC journals had acceptable methodological quality. However, the meta-analytical approach was inappropriate in the majority of the 14 reviews. Some key features of meta-analysis, including computation of appropriate effect sizes, graphic presentation of results on forest plots, assessment of heterogeneity or publication bias, were inadequate in most of the examined reviews. As opposed to multivariate network-analyses or Bayesian approaches most meta-analyses were conducted using simple univariate methods possibly due to a low volume of data and a user-friendly interface of the software packages offering these methods. Since meta-analysis can guide decision and policy making in the clinical practice, standardised guidelines are required to assess the statistical quality of the method.

References

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). Introduction to meta-analysis. UK: Wiley.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. British Medical Journal, 339, b2535.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology, 7(1), 1-7.

3 pm - 3:30 pm

Bianca Annabelle Simonsmeier

“The positive effects of patient education on health: Insights from a second-order meta-analysis.”

Background. Patient education has multiple benefits as compared to surgery or medications, such as barely any complications or side effects and lower costs. This raises the question whether health education can support routine care in medical contexts and whether health education can be used to substitute different practices in medical contexts.

Objectives. Patient education summarized any form of an intervention that has the goal to promote patients health related knowledge, attitudes, or skills. Most commonly, interventions include demonstration and practice of simple health related skills (for example how to properly set an insulin shot) or explanations of health related content (for example outline the negative consequences of a lack of movement after surgery. Other interventions are more complex including different content and approaches. The goal of the present second-order meta-analysis was to summarize the existing evidence on patient education by investigating the overall effectiveness of patient education across a variety of diseases and health outcomes.

Research questions. The study followed three research questions. First, is patient education effective? Second, how consistent is the effectiveness of patient education across different diseases and health outcomes? Third, is the relation between patient education and health outcomes mediated by third variables such as knowledge?

Method. We conducted a systematic literature search in four electronic databases (PsychINFO, PubMed, Cochrane Library, Eric) and an exploratory hand search to identify suitable meta-analyses on patient education. The search revealed 463 meta-analyses, of which 41 could be included in the second-order meta-analysis. The meta-analysis reported 163 relevant effect sizes on patient education. We conducted the statistical synthesis using robust variance estimation (Tanner-Smith, Tipton, & Polanin, 2016), which permits the inclusion of statistically dependent effect size estimates in a single meta-analysis without requiring information about the inter-correlation between effect sizes within studies. Given the presumed heterogeneity, random effects statistical models were used for all analyses.

Results. The mean effect size of patient education on health outcomes was Cohen's d = 0.30, 95% CI [.17, 0.44]. The majority of effect sizes came from randomized controlled trials, the effect can therefore be interpreted as causal. The effect was found for different diseases (based on the ICD-10 classification) and variations in the implementation of the intervention (for example interventions targeting fact knowledge, self-management or attitudes). Patient education was effective to improve physiological outcomes (such as blood sugar), psychological outcomes (such as anxiety), or pain. Further, positive effects of patient education were also found for knowledge, attitudes and skills, which are potential mediators of the relationship between patient education and health outcomes.

Conclusions and implications. Overall, patient education is an effective intervention to improve health outcomes. The effect is rather causal than correlational as demonstrated in studies implementing randomized controlled trials. The effects hold for both acute and

chronical illnesses and can be found for a variety of health outcomes. The results have great implication for medical practices, demonstrating the effectiveness of cost effective intervention with almost no side effects. It is discussed that patient education may be even more effective when implementing instructional design principles and tailoring the intervention to patients of specific age groups, heritage, or educational levels.

3:30 pm - 4 pm

Dr. Katja Matthias, PD Dr. Matthias Perleth

"Science meets Reality - systematic reviews for health policy decisions according to new psychotherapeutic methods in Germany."

Background. The Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) is the highest decision-making body of the joint self-government in Germany [1]. It is comprised of associations of office-based physicians, dentists, hospitals, health insurance funds and patients' representatives. The G-BA issues legally binding directives and thus determines the details of the statutory health regulations. Aside numerous responsibilities the G-BA assesses new diagnostic and therapeutic methods as well as methods already covered by statutory health insurance. The G-BA Rules of Procedure (Verfahrensordnung, VerfO) primarily regulate the methodological requirements for the scientific assessment of the patient benefit, necessity, and cost-effectiveness of measures as a basis for resolutions [2]. The scientific assessment of the benefit of all, including psychotherapeutic methods, follows the criteria of evidence-based medicine. The selection and evaluation of scientific literature is the foundation for decisions on the inclusion or exclusion of medical methods in (from) the statutory health insurance schedule of benefits. Benefit has to be demonstrated along patient relevant outcomes, as recovery, relief from pain or discomfort, improvement in quality of life, reduced mortality, or reduction of side effects. The G-BA compares these results with treatment options that are already available. To prove the patient benefit only systematic reviews of randomised controlled trials (RCTs) or RCTs are acceptable with few exceptions.

Objectives. The G-BA assessment process and the impact of high quality systematic reviews is explained by the case study of the psychotherapeutic method of Eye Movement Desensitization and Reprocessing (EMDR) in outpatient care for Posttraumatic stress disorder (PTSD) in adults [3]. EMDR is a standardized psychotherapeutic treatment method aimed at the processing of events and experiences that have been traumatic. PTSD is a mental disorder that is associated with characteristic and high symptom exposure and high co-morbidity to other mental illnesses. The prevalence of PTSD in trauma patients can reach up to 50% depending on the type of trauma.

In 2011, a joint application of the federal association of health insurance funds and patient

representatives was submitted to consider coverage of EMDR for PTSD.

Research question. In preparation for decision-making the efficacy of EMDR in comparison with unspecific treatment interventions for PTSD (e.g. waiting list, treatment as usual or relaxation methods) or specific interventions (e.g. other psychological treatments already covered by statutory health insurance) was determined using the change of PTSD symptoms by means of standardized instruments as the main outcome measure.

Methods. Comprehensive electronic database searches for RCT or systematic reviews took place in in July 2011 and in April 2013. The database search included the Cochrane Library, PubMed (Medline), EMBASE, PsycInfo and Psyndex. A priori criteria for considering studies for this assessment included participant characteristics, diagnosis, comorbidity, setting, comparators and outcome measures. Two or more members of the working group independently identified studies, assessed trial or review quality and extracted data. Moreover, written statements to the G-BA received from scientific, patients' and other associations as well as from individuals were integrated into the review process.

Findings. After removing duplicates, 1836 records were identified. 1431 records were excluded on the base of information provided in the titles and abstracts. 27 records including nine systematic reviews and 12 RCTs were finally selected. Only two out of nine systematic reviews were of sufficient quality according to critical appraisal. Since these two reviews no longer reflected the current state of research due to the date of their last database searches in 2007 and 2008, an evaluation based on the primary studies only, including meta-analysis, was completed. The final systematic review of the working group included 12 RCTs in total and nine in the meta-analysis. All primary studies included only small sample sizes and most of the studies had a moderate to high risk of bias. There was also great variability between studies with regard to the number of EMDR sessions (1 to 12). In seven out of 10 studies significant improvements in the EMDR-group on clinician rated traumatic stress symptoms or self-reported PTSD symptom severity compared to an unspecific treatment group were seen. The comparison of EMDR with a specific intervention showed no statistically significant differences in five out of six studies. However, the results of the meta-analyses showed a marked benefit of EMDR as compared to standard treatment or unspecific treatment on disease-specific symptom scales post treatment. In 2014, a written hearing process was completed and coverage was concluded in October 2014.

Conclusions and implications. The use of systematic reviews for comparative benefit assessment in the German health care system has been established for many years and contributes to the efficiency of the statutory health care system. The quality of systematic reviews has to be further improved to enable fast health policy decisions.

[1] http://www.english.g-ba.de/

[2] https://www.g-ba.de/informationen/richtlinien/42/ (German only)

[3] https://www.g-ba.de/informationen/beschluesse/2085/ (German only)

4 pm - 4:30 pm

Bianca Annabelle Simonsmeier

"The relationship between subjective well-being and physical activity: A meta-analysis"

1. Background

Subjective well-being (SWB) is a key variable for success across different life domains (Lyubomirski, King, & Diener, 2005) and mental health (Helliwell, 2003). Therefor, it is desirable to promote SWB in the short- and long-term. One possibility to promote SWB is physical activity (e.g., Solberg, Halvari, & Ommundsen, 2013; Mura, Sancassiani, Migliaccio, Collu, & Carta, 2014), as it likely facilitates detachment from work demands, promotes a wide range of psychological needs, and instigates physiological mechanisms. Although the connection between SWB and physical activity and is already part of common knowledge and has been examined numerous times in scientific research, findings sometimes contradict each other or result in different magnitudes of effect sizes. The present study aimed to fill this gap by analyzing the relationship between the two constructs meta-analytically and analyzing potential third variables moderating the relationship.

2. Objectives and Research Questions

It was hypothesized that there is a significant positive relationship between physical activity and SWB for the general population. Previous research solely focused on the link between physical activity and positive affect (Reed & Ones, 2006) or SWB in older adults (Netz, Wu, Becker, & Tenenbaum, 2005), but both with positive small to moderate effect sizes. Additionally, it was hypothesized that the relationship is influenced by third variables, for example exercise duration and intensity or subcomponents of SWB, as found in previous empirical data. The moderator analyses were rather explorative and no directional assumptions were made.

3. Method

Studies were located through the electronic database PsycINFO using a search string that contained relevant key terms. Only studies that observed non-disordered human beings and that were published in English language in peer-reviewed journals were included into the analysis. A total of 837 studies were located and used for further analyses. After coding the abstracts and full-text based on pre-determined inclusion criteria and coding rules, a final

sample of 83 studies remained for conducting the meta-analysis. The included studies reported data from 97 different samples and provided 1290 different effect sizes. The total sample size comprised 467,387 individuals.

We conducted the meta-analysis using robust variance estimation (Tanner-Smith, Tipton, & Polanin, 2016), which permits the inclusion of statistically dependent effect size estimates in a single meta-analysis without requiring information about the inter-correlation between effect sizes within studies. Given the presumed heterogeneity, random effects models were used for all analyses. To address study artifacts that alter the value of outcome measures, we made corrections to the correlations obtained from the single studies for measurement error (Schmidt & Hunter, 2015). We visually and statistically tested for publication bias (Duval & Tweedie, 2000; Egger, Smith, Schneider, & Minder, 1997).

4. Results

The overall correlation between SWB and physical activity was significant with $r = .26$ with a standard error of 0.05. With a certainty of 95% the mean of the distribution of effects sizes lies within the range between .17 and .35. The index $I^2 = 99.72$ indicates a highly heterogeneous pool of effect sizes. Among the different well-being measures, bodily well-being ($r = .32$) and health related quality of life ($r = .33$) showed the highest correlations. The correlations of aerobic activity ($r = .24$) and hard intensity ($r = .32$) were the highest among the exercise characteristics. Further, the relationship between SWB and physical activity was found to increase as people become older.

5. Conclusions and implications

The results underline the positive relationship between SWB and physical activity and provide an insight into the link between body and mind. They might conceivably be useful for helping people get engaged into exercise or for promoting effective exercise programs. While many of the included studies provided evidence for the causal effect of physical activity on SWB, which is yet to be analyzed, less is known about causal effects of well-being components on physical activity, even though SWB is known to increase the frequency of participation in health behaviors, such as physical activity (Diener et al., 2017). Additionally, as participants in this analysis were drawn from the broad healthy majority, the relationship might even be higher for clinical populations.

Session 5: **Applications: Assessment Issues**

Monday, June 11
5 pm - 6:30 pm

5 pm - 5:30 pm	*Jessica Daikeler, Henning Silber, Michael Bosnjak* "Where do web surveys work? A meta-analysis."
5:30 pm - 6 pm	*Caroline Marker, Markus Appel, Timo Gnambs* "Social network site use and academic achievement: Four Meta-Analyses."
6 pm - 6:30 pm	*Sameh Said-Metwaly, Belén Fernández-Castilla, Eva Kyndt, Wim Van den Noortgate* "How Do Testing Conditions Affect Creative Performance? Meta-Analyses of the Effects of Time Limits and Instructions."

5 pm - 5:30 pm
Jessica Daikeler, Henning Silber, Michael Bosnjak
"Where do web surveys work? A meta-analysis."

Background. In an increasingly globalized world cross-cultural research questions and thus cross-national datasets become increasingly important. Against the background of cost-intensive and inflexible face-to-face surveys, there are international attempts for web-based cross-cultural data collections. One of the major challenges in web-based data collection is nonresponse bias. Obviously, the nonresponse rate is not equal with nonresponse bias. Nonetheless, those two concepts are strongly related and moderated by survey design features such as the survey population (Groves & Peytcheva 2008). However, previous research which aimed at explaining web response rates (e.g., Manfreda et al. 2007) could not explain large parts of response rate heterogeneity. The reason for this might be that only survey design factors were included as explanatory variables. Thus, cultural variables such as the structural variables, population-based variables, and variables related to the survey climate could be key variables that might help to explain the heterogeneity across countries. Including cultural variable in explanatory models is line with current cross-national research which showed that cross-cultural differences could affect respondents response behavior (e.g., Stark et. al. 2018, He 2014, Rammstedt 2017).

Objectives and Research question(s) and/or hypothesis/es. In a first step, we explore whether there are cross-cultural differences in web response rates. In a second step, we investigate whether cultural variables can explain web response rate differences between countries.

Method/Approach. We applied a random- effects meta-analytical approach and used

country-aggregated variables. The pool of studies only included studies with an experimental design which compared the web mode to another mode. This resulted 120 studies from 7 countries (Australia, The Netherlands, Sweden, Slovenia, Germany, UK, US).

On the cultural level, we used variables related to a country's structure (internet and phone coverage rates, economic power and progressiveness), variables to countries population (population age, education and cognitive capacities, attitudes such as altruism and collectivism), and variables related to the survey climate (frequency of surveys, privacy concerns).

Results/Findings. Across more than hundred experimental web mode comparisons from seven different countries, we found significant country-based differences for both- the web response rate as well as for the response rate difference between web and the comparison mode. We conclude that national country-based explanatory variables significantly affect a country's mean web response rate and response rate difference.

Conclusions and implications (expected). These findings provide important information for researchers who plan national and cross-national data collections by helping them to evaluate a web surveys response rate in a certain countries before the data collection.

5:30 pm - 6 pm

Caroline Marker, Markus Appel, Timo Gnambs

"Social network site use and academic achievement: Four Meta-Analyses."

Background. Social network sites (SNS) are often blamed for students' poor school performance and bad grades (by educators or parents; e.g., Writer, 2013, but the empirical evidence is all but conclusive. Some studies failed to find any relationships between SNS use and academic achievement (e.g., Hargittai & Hsieh, 2010), while others found negative relationships (e.g., Kirschner & Karpinski, 2010), or even positive relationships (e.g., Khan, Wohn, & Ellison, 2014). Also, theoretical considerations show two possible (conflicting) mechanisms. On the one hand, the time displacement hypothesis (Nie, 2001; Putnam, 2000; cf. Tokunaga, 2016) argues, that the more time a person spends online on SNS, the less time remains for studying, which might result in poorer grades. On the other hand, SNS can help to develop social capital (e.g., Ellison, Steinfeld, & Lampe, 2007; Resnik, 2001), which can be a resource for academic achievement through the help of classmates. Thus, an open question regarding the relationship between SNS use and academic achievement remains, leading to the need for a meta-analytic summary: does

SNS use impair academic achievement?

Objectives. Despite a large number of studies on the relationship between SNS use and academic achievement, the empirical findings are not conclusive. This meta-analysis aimed to identify the association between different types of SNS use and academic achievement. Moreover, a possible explanation offered by the time displacement hypothesis was tested.

Research questions. We distinguished a priori three types of SNS use: (a) a general SNS use including measures like time spent on SNS or frequency of logins; (b) a multitasking SNS use for example in times of studying; (c) a SNS use for academic purposes including activities in learning groups on SNS. For the first two types of SNS use we expected negative relationships while for the third we expected a positive relationship with academic achievement. As an open research question, we tested the time displacement hypothesis to see if more time spent with SNS is associated with less time spend for studying and therefore would reduce academic achievement.

Method. We conducted four random-effects meta-analyses including 59 independent samples (N = 29,337), addressing three different patterns of SNS use, as well as a meta-analytic structural equation analysis. SNS use was differentiated into a general SNS use (55 samples), multitasking SNS use (15 samples), and SNS use for academic purposes (ten samples).

To meet our inclusion criteria the studies had to report (1) a measure of SNS behavior, (2) a measure of school achievement (i.e., school grades), (3) the sample size and a measure of association between the two variables. We excluded studies with general internet measures as well as studies that compared SNS users to non-users or measured attitudes towards SNS use. For academic achievement we excluded measures of cognitive performance and focused only on actual school grades. Two coders were trained and coded all relevant information with an excellent intercoder reliability of Krippendorff's (1970) $\alpha = 1.00$ (based on a subset of 120 effect sizes). We further coded variables for later moderator and sensitivity analyses.

Results. The meta-analyses identified small negative associations of r = -.07, 95% CI [-.12, -.02] for general SNS use and academic achievement, and r = -.10, 95% CI [-.16, -.05] for SNS use related to multitasking and academic achievement. However, SNS use for academic purposes was positively related to academic achievement with r = .08, 95% CI [.02, .14]. We also tested the time displacement hypothesis as a theoretical explanation for the negative relationships. The meta-analytic structural equation analysis showed no support for a time displacement of study time through time spent with SNS. Time spent on SNS was not related to study time, r = -.03, 95% CI [-0.11, 0.06] and, consequently,

showed no mediating effect of academic achievement.

Conclusions and implications. Our meta-analytic summary underscores the notion that SNS use is positively associated with academic achievement as long as SNS use is school-related. This is in contrast to fears of many parents and teachers that the influence of SNS is inevitable detrimental for academic achievement. SNS use unrelated to school, however, was associated with poorer academic achievement. All correlations identified in these meta-analyses were rather weak, only a small part of students' achievement at school and university co-varied with SNS use. A meta-analytic investigation of the time displacement hypothesis found no support for the assumption that the intensity of social media activities is associated with less time spent for studying. Despite the proliferation of SNSs in societies around the world, social networking activities appear to be only weakly related to academic achievement.

6 pm - 6:30 pm

Sameh Said-Metwaly, Belén Fernández-Castilla, Eva Kyndt, Wim Van den Noortgate

"How Do Testing Conditions Affect Creative Performance? Meta-Analyses of the Effects of Time Limits and Instructions."

Background. Over the past decades, a great deal of research has been devoted to the question of whether testing conditions affect performance on creativity tests. Time limits and instructions given to participants have been the subject of many studies investigating their impact on creative performance. However, the results of these studies have been inconsistent. Hence, there is still a lack of clarity regarding the impact of these two factors on creative performance.

Objectives. The purpose of this study is to meta-analyze previous research results regarding the effects of varying time limits and instructions on creative performance and investigate potential moderator variables of these effects.

Research questions. This study attempts to answer the following questions: (1) Do varying time limits (short vs. long) influence creative performance? (2) Do varying instructions (standard vs. explicit instructions to "be creative") influence creative performance? (3) Are there variables that moderate the size of these effects?

Method. Two meta-analyses were conducted; one for time limits effect and one for instructions effect. We also examined how these effects are moderated by measurement approach, gender, culture, education stage, domain of creative performance, and study quality. A meta-analytic three-level model was employed in order to account for dependence within studies.

Results. For time limits, the meta-analysis of 35 effect sizes from nine selected studies

resulted in a relatively large overall effect size for the sake of long time limits. Analysis at the subscale level showed that compared to short time limits, long time limits had a significant positive effect on the originality scores, but no significant effect on either fluency or flexibility. None of the moderator effects attained statistical significance at either the whole or subscale level, but this could be due to the low power of these analyses.

Regarding the meta-analysis of the effect of the type of instructions, 92 effect sizes obtained from 30 studies were analyzed, yielding a non-significant overall effect size. An investigation at the level of the subscales showed that explicit instructions to "be creative" enhanced originality scores in comparison to standard instructions, with no significant effects on either fluency or flexibility. Only education stage significantly moderate the effects of instructions on both originality and fluency. Compared to college samples, non-college samples were found to have a significantly larger mean effect size for originality and a smaller mean effect size for fluency.

Conclusions and implications. The results of this study suggest that time limits and instructions given to subjects could significantly influence their creative performance. Our results inform the ongoing debate concerning the optimal conditions for administering creativity instruments, and have valuable implications for researchers and educators interested in measuring creativity. The evidence from this study indicates that giving subjects longer time limits and explicit instructions to "be creative" when testing creativity could result in greater originality and overall scores. Also, caution should be exercised when comparing the results obtained from studies with different testing conditions.

Session 6: **Methodological Issues (Part 2)**

Tuesday, June 12
10:30 am - 12 pm

10:30 am - 10:50 am	*Caspar J. van Lissa* "MetaForest: Exploring heterogeneity in meta-analysis using weighted random forests."
10:50 am - 11:10 am	*Michael Kossmeier, Ulrich S. Tran, Martin Voracek* "Visual inference for the funnel plot in meta-analysis."
11:20 am - 11:40 am	*Tessa van den Berg, Suzanne Jak* "Comparing Meta-Analytic Structural Equation Modeling and Univariate Meta-Analysis: An Application in Forensic Child and Youth Care Sciences."
11:40 am - 12 pm	*Peter Schmidt* "A Meta-Analytic Structural Equation Model for the Theory of planned Behavior."

10:30 am - 10:50 am
Caspar J. van Lissa
"MetaForest: Exploring heterogeneity in meta-analysis using weighted random forests."

Background. Meta-analysis is the practice of synthesizing evidence from multiple studies by computing a summary effect size (Braver, Thoemmes, & Rosenthal, 2014; Laws, 2016). One requirement of meta-analysis is that the studies being aggregated are conceptually similar, and ideally, close replications (Fabrigar & Wegener, 2016; Higgins, Thompson, & Spiegelhalter, 2009; Maxwell et al., 2015). However, in psychology and other fields, it is common practice to aggregate studies which have investigated similar research questions in different laboratories, using different methods, instruments, and samples. These differences between studies can introduce substantial heterogeneity between studies. Three approaches have been proposed to deal with between-studies heterogeneity (Higgins et al., 2009): First, if studies are assumed to be different, they should not be meta-analyzed. Secondly, if they are similar, a random-effects model can estimate the distribution of the true effect size. Thirdly, if known differences between studies introduce heterogeneity, these moderators can be accounted for using meta-regression. However, the number of studies on any topic is typically low, because research is cost- and time-intensive. Therefore, meta-regression often lacks the power to adequately account for between-studies heterogeneity. Moreover, there is typically a paucity of theory regarding the sources of between-studies

heterogeneity, which could help whittle the number of potential moderators down to a manageable number (Thompson & Higgins, 2002). Heterogeneity between studies thus presents a non-trivial challenge to data aggregation using classic meta-analytic methods (Riley, Higgins, & Deeks, 2011). At the same time, it also offers an unexploited opportunity to learn which differences between studies have an impact on the effect size found. What is currently lacking is a "fourth approach" for dealing with heterogeneity: An exploratory technique which can perform variable selection, identifying which moderators most strongly influence the observed effect size, even when the number of moderators is relatively large.

Objectives. MetaForest aims to address this need. This technique applies random-effects weights from classic meta-analysis to random forests' bootstrapping procedure. Random forests are a powerful learning algorithm, flexible yet relatively robust to overfitting.

Research question. Can MetaForest be used to explore heterogeneity in meta-analytic data, even when the number of cases is low, relative to the number of moderators?

Method. A simulation study was conducted, with five design factors: Number of studies k (20, 40, 80, 120), average within-study sample size n (40, 80, and 160), number of "distractor" moderators M (1, 2, and 5), population effect size of "true" moderators (.2, .5, and .8), and residual heterogeneity (based on a recent open-data study of 705 published meta-analyses, we selected the values 0, .04, and .28, to cover a range of residual heterogeneity values as encountered in real data). Data were simulated using the random-effects model, based on five linear models: (a) main effect of one moderator, (b) two-way interaction, (c) three-way interaction (d) two two-way interactions, (e) non-linear, cubic relationship. Performance was evaluated in terms of three metrics: 1) Predictive performance on new data; 2) power, as evidenced by the proportion of datasets in which the algorithm achieved a positive R2 in a validation dataset; and 3) the ability to distinguish relevant moderators from irrelevant moderators, using variable importance measures.

Results. We found that, even in datasets as small as 20 cases, MetaForest had excellent performance on all three metrics, except when effect size was small, and residual heterogeneity was high. Performance decreased slightly when the number of irrelevant moderators increased. This is because, in the presence of many irrelevant moderators, the random forests algorithm may be forced to choose amongst only irrelevant moderators at some splits. This problem is mitigated by the fact that the out-of-bag R2 indicates when the algorithm is overfitting noise (i.e., it will become negative). We also found that unweighted MetaForest (normal random forests) outperformed random-effects weighted MetaForest when the effect size was large, and when the model included only one relevant moderator. This might be caused by an overestimation of the residual variance: Because MetaForest estimates residual heterogeneity from the raw data, without accounting for the influence of the moderators, it will be overestimated when the amount of heterogeneity introduced by

moderators is relatively large, compared to the residual heterogeneity.

Conclusion and implications. MetaForest is a powerful tool for exploring heterogeneity in meta-analysis. It can identify important moderators from a larger set of potential candidates, even when the number of studies is low. If moderators are continuously distributed, MetaForest often has sufficient power with as little as 20 studies. This is an appealing quality, because many meta-analyses have small samples. MetaForest yields variable importance metrics, which can be used to identify important moderators, and offers partial prediction plots to explore the shape of the marginal relationship between moderators and effect size. Although MetaForest constitutes a fully-fledged paradigm for data analysis, it can also be readily integrated in classical meta-analyses, to ensure that important moderators have not been overlooked (Curry et al., in press). The technique is available as an R package ("metaforest"), and online (www.developmentaldatascience.org/metaforest). Complete results of the simulation study are available on the OSF, https://osf.io/khjgb/.

10:50 am - 11:10 am

Michael Kossmeier, Ulrich S. Tran, Martin Voracek

"Visual inference for the funnel plot in meta-analysis."

Background and Objectives. The funnel plot is a widely used diagnostic plot in meta-analysis to assess small-study effects and publication bias in particular (Light & Pillemer, 1984). The funnel plot was one of the first proposed genuine plots to visualize meta-analytic data and is, next to the forest plot (Lewis & Clarke, 2001), the most iconic and popular display for this purpose (Schild & Voracek, 2013).However, despite the popularity of the funnel plot, its suitability to detect publication bias has been questioned in the past (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). Experimental research showed that conclusions and interpretations based on the visual inspection of funnel plots are subjective and often wrong regarding the presence or absence of publication bias (Terrin, Schmid, & Lau, 2005). Formal statistical tests (e.g., Egger's regression test) based on the asymmetry of the funnel plot are widely used to establish objectivity in the assessment of small-study effects, while controlling for the error of the first kind. All these tests have in common that they are entirely based on funnel-plot asymmetry quantified via the association between (a function of) study effect sizes and (a function of) study standard errors. Visual inspection of the funnel plot on the other hand allows to exploratory incorporate a multitude of visually displayed statistical information to assess the presence and severity of publication bias. For instance: Which role does statistical significance play (using significance contours)? Is there an abundance of just-significant results? Is asymmetry driven by single outlying studies or clusters of studies? As prominently

outlined in the Cochrane handbook, formal tests for funnel-plot asymmetry should never be interpreted in isolation, but rather always in the light of visually inspecting the funnel plot (Sterne, Egger, & Moher, 2008, p. 317).Hence, visual inference might be the sought-after bridge between both worlds: It allows researchers to formally guard against type I errors, while still keeping the more general diagnostic, open, and explorative nature of examining funnel plots visually.

Methods. Visual inference is a formal inferential framework proposed by Buja et al. (2009) that allows to test whether graphically displayed data do or do not support a hypothesis. The principal idea is that if a suitable statistical plot of the observed data is visually distinguishable from plots of data simulated under the null hypothesis, then this constitutes evidence against the null hypothesis.

The central procedure to draw valid inferences is carried out with the so-called lineup protocol: A lineup of diagnostic plots is constructed for observed data and a null hypothesis a researcher wants to reject. The lineup consists in total of k plots, with k-1 plots showing data simulated under the null hypothesis and one plot showing the actually observed data at a randomly chosen position. The lineup is then presented to a viewer unfamiliar with the observed data and their peculiarities. If the plot showing the actually observed data is noticeably different and therefore identifiable by the viewer out of all plots in the lineup, then the null hypothesis is rejected.

If the observed data in fact are realizations of the null hypothesis, the probability to identify the plot showing the real data and therefore to falsely reject the null hypothesis is 1 divided by k, the number of plots in the lineup. The alpha level is therefore controlled by the size of the lineup. The most natural choice for the number of plots in the lineup is 20, corresponding to the conventional alpha level of 5%. Just like for conventional statistical tests, a test statistic (the actual plot) is compared to the null distribution (the plots showing data simulated under the null hypothesis) and assessed whether it is plausible or extreme. The difference is that the test statistic is not compared to the whole null distribution, but rather to a finite number of realizations of this distribution (Majumder, Hofmann, & Cook, 2013).

Results. We suggest funnel plots as a prime candidate field for the application of visual inference for two main reasons. First, visual inference has the potential to increase the (often low) validity of funnel plot based conclusions by controlling the error of the first kind. The lineup protocol allows researchers to guard themselves of prematurely interpreting patterns in the funnel plot which might indeed be perfectly plausible by chance. Second, at the same time visual inference allows using visual perception as a formal statistical test and therefore allows to flexibly incorporate a mutilate of visual information. Conventional funnel plot based statistical tests to assess small study effects, e.g., Egger's regression test, exclusively focus on the association of (a function of) effect sizes and standard error, while visual inference remains the explorative nature of diagnostic

graph inspection.

For practitioners of meta-analysis, an important question is how to conveniently conduct visual funnel plot inference in practice. Within the statistical computing environment R (R Core Team, 2017), the package nullabor (Wickham, Chowdhury, & Cook, 2014) is available, which provides helpful general-purpose functions to conduct visual inference with arbitrary graphical displays. Building on this, we develop and document the R function funnelinf to specifically conduct visual inference with funnel plots. This function depends on nullabor, but provides features tailored for visual inference with funnel plots. Current key features of the function are: (1) options for null-plot simulation under both classic meta-analytic models (fixed-effect model, random-effects model), (2) subgroup analysis, (3) different funnel-plot specific graphical options (significance and confidence contours, choice of y-axis), and (4) options to display additional statistical information (Egger's regression line, imputed studies, and adjusted summary effect by the trim-and-fill method).Visual funnel-plot inference improves the validity of conclusions based on visually inspecting the funnel plot by controlling the error of the first kind. Despite this important key feature, a natural question is how often the procedure leads to the correct rejection of the null hypothesis, if in fact there is true signal in the data. We address this research question in a pilot experiment study, by investigating the power of visual funnel-plot inference to detect (simulated) publication bias in different scenarios.

Conclusions and Implications. We propose to present a funnel plot of the actually observed data simultaneously with null funnel plots, showing data simulated under some suitable null hypothesis. Only if the funnel plot showing the real data is identifiable out of all plots, the null hypothesis is formally rejected and conclusions based on visually inspecting the funnel plot might be warranted. We recommend that visual funnel-plot inference should be used routinely, as it is a convenient way to increase the validity of conclusions based on funnel plots by guarding the investigator from interpreting patterns in the funnel plot that might be perfectly plausible by chance. Software to conduct visual inference with funnel plots is provided in the form of a tailored R function.

References

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E. K., Swayne, D. F., & Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 367, 4361-4383.

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. British Medical Journal, 333, 597.

Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. British Medical Journal, 322, 1479–1480.Light, R. J., & Pillemer, D. B. (1984). Summing up: The

science of reviewing research. Cambridge, MA: Harvard University Press.

Majumder, M., Hofmann, H., & Cook, D. (2013). Validation of visual statistical inference, applied to linear models. Journal of the American Statistical Association, 108, 942-956

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Schild, A. H., & Voracek, M. (2013). Less is less: A systematic review of graph use in meta-analyses. Research Synthesis Methods, 4, 209–219.

Sterne, J. A., Egger, M., & Moher, D. (2008). Adressing reporting bias. In J. P. Higgins, & S. Green (Eds.), Cochrane handbook for systematic reviews of interventions (pp. 297-333). Chichester, England: Wiley.

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. Journal of clinical epidemiology, 58, 894-901.

Wickham, H., Chowdhury, N. R., & Cook, D. (2014). Nullabor: Tools for Graphical Inference. R package version 0.3.1. https://CRAN.R-project.org/package=nullabor.

11:20 am - 11:40 am

Tessa van den Berg, Suzanne Jak

"Comparing Meta-Analytic Structural Equation Modeling and Univariate Meta-Analysis: An Application in Forensic Child and Youth Care Sciences."

Background. Meta-analysis is a widely used statistical technique to integrate research findings of a large collection of studies, which results in a summary effect size (Glass, 1976). The most used approach of meta-analysis, univariate meta-analysis, tests a hypothesis about one association between two variables. However, research questions often involve more complex relations, such as mediation effects. These research questions cannot be answered using univariate meta-analysis. Meta-analytic structural equation modeling (MASEM, Viswesvaran & Ones, 1995), combining meta-analysis and structural equation modeling, can answer these questions by testing entire models of sets of variables at once.

Objectives. Since MASEM is a relatively new approach, researchers may not be acquainted with and not be aware of the advantages and disadvantages of MASEM. This study compares MASEM to univariate meta-analysis, and illustrates that choosing one of these research methods over the other could have an important impact on the outcomes. By using illustrative data, the different characteristics of the two methods become clear. The explanation of the advantages and disadvantages will help researchers to choose the most

appropriate method for their research.

Research Question(s). This study illustrates the differences between MASEM and univariate meta-analysis, and discusses the advantages and disadvantages of both methods.

Method/Approach. The characteristics of both methods were illustrated by the research project 'The potential mediating role of parenting on intergenerational continuity of criminal behavior'. The hypothesized model implies that the effect of parental delinquency on juvenile delinquency is partially mediated by several parenting behaviors (support, authoritarian control, behavioral control, psychological control, and indirect parenting). A total of 88 studies were included in the study (total $N = 154{,}176$). In the univariate meta-analysis, pooled correlations were estimated for each association in the hypothesized model. MASEM was conducted using the Two Stage SEM approach (Cheung, 2014; Jak, 2015): a pooled correlation matrix was estimated in the first stage, on which the hypothesized path model was fitted in the second stage.

Results/Findings. Effects from parental delinquency to parenting were almost similar and the same conclusions can be drawn based on the two methods. Small differences arose due to the fact that MASEM is a multivariate analysis and takes into account that some effect sizes depend on each other because they came from the same study, while univariate meta-analysis does not. The effects from parenting to juvenile delinquency and from parental delinquency to juvenile delinquency showed remarkable differences. All parameter estimates (multiple regression coefficients) in MASEM were smaller than the pooled correlations in univariate meta-analysis. Importantly, this leads to different conclusions about the research questions. In particular the relationships between psychological control and juvenile delinquency, and between indirect parenting and juvenile delinquency showed large differences. Both parameter estimates were not significant in MASEM ($\beta = -0.011$, 95% CI [−0.102, 0.080] and $\beta = -0.093$, 95% CI [−0.186, 0.001] respectively), while significant correlations were obtained in univariate meta-analysis ($r = 0.079$, 95% CI [0.020, 0.137] and $r = -0.178$, 95% CI [−0.251, −0.105] respectively). In univariate meta-analysis, indirect parenting even seems to have the largest influence on juvenile delinquency, compared to other parenting variables. Apparently, the effect of these parenting variables on juvenile delinquency disappears when controlling for other parenting variables and parental delinquency.

The moderator analysis of juvenile's age in univariate meta-analysis and MASEM resulted in different outcomes as well. Since MASEM uses subgroup analysis to investigate moderators (Jak & Cheung, under review), this continuous moderator variable had to be categorized for the analysis in MASEM, leading to a loss of information. As a consequence, MASEM has less power to detect moderating effects than univariate meta-analysis. Indeed, the significant effects that resulted from univariate meta-analysis were not found with MASEM. However, the possible moderator socio-economic status, which was investigated as a categorical variable in both univariate meta-analysis and

MASEM, showed similar results across methods.

Conclusions and Implications. Altogether, the differences in outcomes showed that both methods have their own characteristics which can be an advantage or disadvantage. MASEM seems to examine relationships between variables more accurately. It gives a more elaborate and precise examination of reality than univariate meta-analysis, because it can test hypothesized models, control for other variables, and estimate mediation effects. Moreover, MASEM can investigate research questions that were not explored before, since not every primary study needs to include all variables of the hypothesized model (Viswesvaran & Ones, 1995). The combination of different studies about different associations may give new insights. On the other hand, univariate meta-analysis is more precise when it comes to moderator analysis of continuous variables. The main disadvantage of MASEM is that it can investigate moderators only as categorical moderators. Hence, an important suggestion for future research would be to improve the moderator analysis in MASEM in a way that it can investigate continuous moderator variables as well.

It is very important to keep in mind that differences in effect sizes, as were shown in this study between univariate meta-analysis and MASEM, lead to differences in conclusions about the research questions. This affects everything that depends on these conclusions, such as clinical implications. Therefore, it is very important to choose the most appropriate research method to investigate specifically your research question. Although both methods have their advantages and disadvantages, in social sciences it may be preferred to take several variables into account – since behavior of individuals will always be affected by more than one aspect – and use MASEM as a research method.

References

Cheung, M. W. L. (2014). Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46*(1), 29-40.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3-8. doi:10.3102/0013189X005010003

Jak, S. (2015). *Meta-Analytic Structural Equation Modeling*. Springer International Publishing. doi:10.1007/978-3-319-27174-3

Jak, S., & Cheung, M. W.-L. (under review). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis.

Viswesvaran, C., & Ones, D. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865-885. doi:10.1111/j.1744-6570.1995.tb01784.x

11:40 am - 12 pm

Peter Schmidt

“A Meta-analytic Structural Equation Model for the Theory of planned Behavior.”

a) Background

The issue of insufficient reproducibility of empirical findings in psychology has become a central issue (Science ,349, 2015). One rare exception is the Theory of Planned Behavior. It is one of the most often tested theories in social psychology. Applications include consumer behavior, travel mode choice, health-related behavior, entrepreneurship research, environmentally related behavior and political behavior (Fishbein/Aizen 2010). A lot of conventional meta-analyses have

been performed over different behavioral domains, within behavioral domains and very few for interventions based on the theory(Steinmetz et al. 2016). Most of them used fixed effects models, did not correct for measurement error and did not test sufficiently for relevant moderators like student samples versus population samples and used only correlations as input and output. However the new approach of meta-analytical structural equation modeling (for a summary Cheung 2015) allows to take into account both measurement error and estimates the partialized regression coefficients instead of correlations between constructs as an outcome. Furthermore it allows to test relevant moderators as a multi-group or multilevel model.

b) Objectives

The goal of the paper is firstly to show how partial regression coefficients corrected for measurement error can be estimated based on a meta-analytical structural equation, which can be used as a-priori information in testing the theory with new data-sets.

Secondly it should provide social scientists with knowledge how much moderating effects like type of sample, mode effects, method effects influence and modify the coefficients and have to be taken into account in specifying the models to be tested and in the interpretation of the results.

Thirdly it should demonstrate all this for the theory of planned behavior and provide more realistic apriori assumptions for future studies employing this theory taking into account the varying methodological and contextual conditions of them .

Fourthly it will show how meta-analysis is not only used for calculating mean correlations or regression coefficients but provides information how to clarify unsettled issues in a theory and even modify it.

Fifthly it will demonstrate how the newly developed package Metasem for meta-analytical structural equation modeling on the basis of R by M.Cheung(2015) is used.

c) Research Questions

One central research question is how far our partial regression coefficients based on meta-analytical SEM differ from those in earlier meta-analyses and the correlations mostly used in former meta-analyses. A second research question is to test systematically the effect of different moderators like type of sample, mode of data collection, self-report vs. observation, composite scores or measurement models with multiple indicators etc. For all moderators we set up specific hypotheses to be tested like H 1: For student samples the correlations between all variables involved are higher than for population samples. Another example of a hypothesis is H2 : The correlation between self- report variables is higher than the correlations between observation variables(e.g. behavior) and self report variables(e.g. attitude toward the behavior, norms and perceived behavioral control).

d) Method

Based on 369 studies employing the Theory of Planned Behavior we used 200 correlation matrices for the meta-analytical structural equation model. The sample size of the studies ranged from 23 as minimum to 1797 as maximum. They varied considerably between the different relations between the constructs of the theory of planned behavior. They refer to the following behavioral domains: nutrition, Health in general, Achievement, Sports, Addiction, Use of technical tools, Environment, Leisure Time and Travel Mode Choice. To visualize and test possible publication bias, we employed funnel plots. Cheung (20o8, 2013 d), 2015) demonstrated how meta-analyses can be formalized and tested as structural equation models. For the meta-analysis we used the two-stage structural equation model with random effects (Cheung 2013a, d, 2015). In the first stage one performs a synthesis of the primary correlations whereas during the second step one estimates the structural equation coefficients based on the mean correlations. We have used the R-program developed by M. Cheung for this purpose.

e) Results

The search procedure for the documents, the selection criteria and the coding of the documents will be described in detail according to the standards set by APA. We report in detail the number of correlations, the sample sizes, Means, Medians, Minimum and Maximum Values and Standard Deviations for all relationships like Intention and Behavior, Attitude and Behavior etc. separately as there are considerable differences. Furthermore we used stem and leaf graphs to visualize the variation of the correlations between the constructs. Corresponding to the first step of the two step procedure of meta-analytical structural equation modeling we present a multivariate random effect model based on all correlations. As an output we get 1o mean correlations between the five constructs of the core model of the TOPB. They are-as expected- all highly significant. The multivariate Q test of homogeneity showed highly significant results, which implies strong heterogeneity and the existence of moderator effects. Following this we present the results of the meta-analytical structural equation model based on the mean correlations from step 1 and their asymptotic covariance matrix. As one can expect that the relations may be very different according to the domain of behavior, we did the same two steps for each of the six domains of behavior for which enough studies were available and present the results for the test of homogeneity and the estimates for the domain specific meta-analytic structural equation models.

Session 7: **Applications: Educational & Developmental Psychology (Part 1)**

Tuesday, June 12
2:30 pm - 4 pm

2:30 pm - 3 pm	*Saraswathi Ramasamy* "The formative approach: A conceptualization and a meta-analysis on the effects on learning."
3 pm - 3:30 pm	*Lena Weyers, Martina Zemp, Georg W. Alpers* "Impaired interparental relationships in families of children with ADHD: A meta-analysis."
3:30 pm - 4 pm	*Carolin Schuster, Lisa Pinkowski, Daniel Fischer* "To what extent do individual values change in adulthood? A systematic literature review of longitudinal studies assessing Schwartz' values among adults."

2:30 pm - 3 pm

Saraswathi Ramasamy

"The formative approach: A conceptualization and a meta-analysis on the effects on learning."

Background. The term "formative approach" (FA) characterizes teachers' practices of assessing students' learning progress, providing them with comprehensive feedback and adapting instruction based on the assessment outcomes. FA is widely implemented for its assumed benefits on learning. Previous attempts to understand FA and its impact on learning outcomes by means of meta-analyses have yielded diverse and inconclusive results.

Objective. The goal is to conceptualize formative approach precisely and to determine the impact of the different components of formative approach on learning.

Research question. Therefore, we propose a model of FA and test whether the model is valid. The FA model essentially includes three components, namely, diagnostic assessment, comprehensive feedback and adaptive instruction. The components operate partially in a cumulative manner with the diagnostic assessment component always preceding comprehensive feedback and adaptive instruction.

Method. In the present study, we chose a meta-analytical approach to gather empirical evidence for the proposed model. From an initial search result of 5096 articles on formative practices, we screened 838 empirical and quantitative studies. After applying the inclusion-exclusion criteria we had 100 articles for the analysis. From the 100 articles, we obtained $k = 348$ effect sizes for the analysis. We used dummy coding for the proposed

components of the FA model. Because of the heterogeneity in the conceptualizations of FA between studies, we opted for a random-effects meta-regression model. The analyses were run with the *'metafor'* package in R.

Results. Across all studies, the effect of FA on learning was $d = 0.48$ [0.31, 0.66]. However, an inspection of the studies showed that the conception of FA differed across the studies. Coding showed that 51 of the effect sizes resulted from studies which confined FA only to the diagnostic assessment component, 86 resulted from studies with diagnostic assessment and comprehensive feedback components, 54 resulted from studies with diagnostic assessment and adaptive instruction components and 157 effect sizes resulted from studies with all three components. These differences were associated with differences in the effect sizes. The effect of FA in terms of diagnostic assessment was $d = 1.07$ [0.75, 1.38] whereas it was $d = .66$ [0.35, 0.96] for studies which included diagnostic assessment and adaptive instruction.

Conclusions. The results illustrate that part of the heterogeneity of the findings on the effects of FA on learning result from the diverse understanding of formative approach. The current findings therefore can be considered as a first meta-analytic support for the suggested model specifying the components of FA. Results on potential additional moderators will also be presented and discussed.

3 pm - 3:30 pm

Lena Weyers, Martina Zemp, Georg W. Alpers

"Impaired interparental relationships in families of children with ADHD: A meta-analysis."

Research on ADHD in children and adolescents has traditionally focused most on the genetic and neurobiological aspects of the disorder, but the role of family relationships has been much less systematically examined. There is growing evidence that the quality of interparental relationships and children's ADHD symptoms are reciprocally related. Because previous findings appear to be inconsistent, this meta-analysis aims at summarizing previous research in order to assess whether there are robust differences in the quality of interparental relationships between parents of children with ADHD and parents of healthy children. Fourteen studies with 42 effect sizes supported a small, but significant difference (d = .24) and indicated that parents of a child with ADHD report poorer relationship quality than parents of healthy children. This effect was moderated by the child's age and did not depended on whether the child had a comorbid Oppositional Defiant Disorder or Conduct Disorder.

Keywords: child ADHD; parents, couples; interparental conflict; relationship satisfaction

3:30 pm - 4 pm

Carolin Schuster, Lisa Pinkowski, Daniel Fischer

"To what extent do individual values change in adulthood? A systematic literature review of longitudinal studies assessing Schwartz' values among adults."

Background. Individual human values represent "guiding principles in people's lives" (Schwartz & Bardi, 2001), and predict various socially relevant outcomes, for example voting behavior (Vecchione et al., 2013), pro-environmental behavior (Thogersen & Ölander, 2002), or charismatic leadership (Sosik, 2005). Therefore, the question of whether values change over time–by themselves or through external influences–is not only a basic empirical question, but has further implications for research and interventions aimed at influencing these behaviors.

On the one hand, individual values are commonly considered as being relatively stable over time, especially among adults. On the other hand, it has been argued that values are likely to change to a certain degree (Rokeach, 1973; Schwartz, 1992; Schwartz & Bardi, 1997). However, since the stability of values on the individual level has largely been neglected in psychology (Bardi & Goodwin, 2011; Bardi et al., 2009; Trommsdorff, 1996), it remains unclear under what conditions and to what extent values do actually change in adulthood.

Objectives and Research Questions. The objective of the present review is to provide a comprehensive overview of the current state of knowledge with regard to the changeability of individual general values in adulthood based on the empirical evidence available today. For better comparability and theoretical integration, we focus on Schwartz' value model (Schwartz, 1992), which constitutes a broad and fairly universal set of ten different values which are arranged on a circular pattern according to their compatibility with each other (e.g. *Universalism, Tradition, Power* or *Stimulation*). It is the most commonly applied value concept (Borg et al., 2015) and there are several validated measures for it (Schwartz et al., 2001). Based on the empirical literature, we qualitatively examine the stability and changes of value measurements over time. Given that there is no comprehensive review on this topic yet, it is unclear what kind of maturational, circumstantial or deliberate external influences on values have been examined at all, and the extent to which they elicit changes in a person's value system. On the grounds of the results, we discuss implications for further research on value change. Based on published empirical studies that have examined values over time, the following main research questions are investigated.

1. What is the state of knowledge about the stability of values over time in adulthood?
2. What influences (i.e., things that happen between the measurements) on values have been studied and what is the result?

Method/ Approach. A systematic literature review (Fink, 2014) was conducted to investigate the research questions. In order to identify existing empirical longitudinal studies measuring values among adults, two databases, Scopus and PsycINFO, were used. Different search strings[1] were developed to produce a maximum of relevant, and a minimum of irrelevant papers. We included all peer-reviewed research articles in the subject area of psychology, published in English or German language, which present empirical studies including at least two measurements points of general values based on the Schwartz value model, on a sample of participants of an age of 18 years or older. The initial sample of 273 papers resulting from the search was screened on the basis of these criteria, resulting in 16 relevant publications. By conducting a reference and citation searching on the basis of the pre-final sample the search was extended in two directions, which resulted in 140 additional publications, 1 of which was included in the sample after screening. To validate the results of the literature search, five major experts in the field of human values reviewed the resulting final sample of identified papers and suggested 4 further publications, 1 of which met the inclusion criteria. This process resulted in a final sample of 18 articles. The final papers were then analyzed according to the studies' methodological characteristics and their empirical findings.

Results/ Findings. The identified 18 relevant articles included a total of 27 reported studies, all published within the last two decades. These articles included nine experimental studies from five articles testing interventions to change values (i.e., through priming of values; identification with other people's values; internal consistency maintenance among values and self-perception, moral convictions, strong attitudes, or values-associated feelings; and/ or persuasive approaches). Fourteen studies from ten articles observed value change without intervention, but in the context of certain changes of circumstances (i.e., a national parliamentary election, educational training, higher education, migration, and experiences of war). Five studies from four articles observed stability or changes of values in a more general sample over the course of 1 – 8 years.

The studies further applied different methods to relate the (at least two) measurement points to each other: One way is to examine *mean-level changes* in values (e.g., comparing the mean importance of Universalism in the sample between time 1 and 2). A second way is to examine *rank-order correlations* for each value within a sample (e.g., correlating participants' relative importance of universalism in a sample at time 1 with their relative importance of Universalism at time 2). A third, less frequently used approach is to examine the *stability of value profiles*, that is the within-person correlations of value rankings (e.g., correlating person X's profile at time 1 with his/her profile at time 2).

Overall, the identified studies revealed that basic human values are relatively stable in general. Most rank-order stabilities accounted for correlation coefficients of over .50. However, significant mean-level changes in at least some values were observed in most studies. In the experimental studies, interventions that aimed to strengthen or change

particular values on the mean level were successful. The change in values by such short, one-time interventions appeared to diminish over time, indicating a strong reinforcement of original values in daily routines. Observational studies differed in the total number and types of changing values as well as the direction of change. Under the influence of particular external conditions, expected or at least reasonable value change was observed in most studies. In some cases, however, values did not change as hypothesized or even changed in an opposite direction. The extent of life-changing events, in some of these studies, predicted overall value change. Under general circumstances, the number of changing values differed fundamentally among studies, with more values changing in long-term studies. Indications for a growing value stability by increasing age, and a particular direction of value change (i.e., increase in Self-transcendence) by maturation were found in the eight-year study. Little can be said about value stability and change on the individual level. The three observational studies that assessed the consistency of individuals' value profiles revealed on average considerable intra-individual stability. Still, high variance of value profile correlations indicated that the intra-individual stability of values over time strongly varied between persons. There are only weak insights about potential causes of value change in non-experimental conditions. The studies' empirical findings suggest that values changed as a reaction to frequent priming, adaptation to new environmental conditions or role expectations, or/ and identification with peers.

Conclusions/ Implications. The identified empirical studies confirm the stable nature of values across different adult samples and settings, but also point to several forms of value change: With increasing age, people tend to value self-transcendence more according to one study; external situational influences sometimes have an effect on values, yet it is not clear why and when; and values can intentionally be changed with experimental interventions. However, the paucity of studies dealing with value stability and change in adulthood and the limited comparability of the studies' findings raise the need for further empirical research: Firstly, with regard to value change interventions, further experimental research is needed to examine the stability of such changes and their impact on attitudes and behaviors. Secondly, further research on value change in the context of external events and experiences is needed to better understand the currently very mixed results. Specifically, we suggest that future studies exploring changes in reaction to different external circumstances build their designs on a common theoretical model of value change (e.g., Bardi & Goodwin, 2011), and focus on the underlying processes. Thirdly, long-term high quality studies with at least three measurement waves are required to gain insights into the development of value stability, to distinguish real change from measurement error, as well as to disentangle the effects of external influences, of maturation, and individual differences. In addition, we suggest that further research examines the stability of intra-individual value profiles, besides rank-order stability and mean-level change measures, and includes in their samples older adults.

References

Bardi, A., & Goodwin, R. (2011). The dual route to value change: Individual processes and cultural moderators. Journal of Cross-Cultural Psychology, 42(2), 271–287.

Bardi, A., Lee, J. A., Hofmann-Towfigh, N., & Soutar, G. (2009). The structure of intraindividual value change. Journal of Personality and Social Psychology, 97(5), 913–929.

Borg, I., Bardi, A., & Schwartz, S. H. (2015). Does the value circle exist within persons or only across persons? Journal of personality, 85(2), 151-162.

Fink, A. (2014). Conducting research literature reviews: From the internet to paper (4th ed.). Los Angeles: Sage.

Rokeach, M. (1973). The nature of human values. New York: Free Press.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology, 25, 1–65.

Schwartz, S. H., & Bardi, A. (1997). Influences of adaptation to communist rule on value priorities in Eastern Europe. Political Psychology, 18(2), 385–410.

Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures: Taking a similarities perspective. Journal of Cross-Cultural Psychology, 32(3), 268–290.

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. Journal of cross-cultural psychology, 32(5), 519-542.

Sosik, J. J. (2005). The role of personal values in the charismatic leadership of corporate managers: A model and preliminary field study. The Leadership Quarterly, 16(2), 221-244.

Thogersen, J., & Ölander, F. (2002). Human values and the emergence of a sustainable consumption pattern: A panel study. Journal of Economic Psychology, 23, 605–630.

Trommsdorff, G. (1996). Werte und Wertewandel im kulturellen Kontext aus psychologischer Sicht. In E. Janssen, U. Möhwald, & H. D. Ölschleger (Eds.), Monographien aus dem Deutschen Institut für Japanstudien / Deutsches Institut für Japanstudien. Gesellschaften im Umbruch? Aspekte des Wertewandels in Deutschland, Japan und Osteuropa (pp. 13–40). München: Iudicium-Verlag.

Vecchione, M., Caprara, G., Dentale, F., & Schwartz, S. H. (2013). Voting and values: Reciprocal effects over time. Political Psychology, 34(4), 465–485.

[1] based on the terms *value change*, *change of values*, *influence on values* and *stability of values*

Session 8: **Applications: Educational & Developmental Psychology (Part 2)**

Tuesday, June 12
4:30 pm - 6:30 pm

4:30 pm - 5 pm	*Hannelies de Jonge, Suzanne Jak* "A Meta-Meta-Analysis: Identifying Typical Conditions of Meta-Analyses in Educational Research."
5 pm - 5:30 pm	*Michael Schneider, Franzis Preckel* "Instruction- and student-related correlates of achievement in higher education: A systematic review of meta-analyses."
5:30 pm - 6 pm	*Bianca Annabelle Simonsmeier* "Domain-Specific Prior Knowledge and Learning: A Meta-Analysis."
6 pm - 6:30 pm	*Peter Edelsbrunner, Christian M. Thurn* "The Prevalence of Unfounded (Statistical) Inferences Based on non-significant p-values in Educational Psychology."

4:30 pm - 5 pm
Hannelies de Jonge, Suzanne Jak
"A Meta-Meta-Analysis: Identifying Typical Conditions of Meta-Analyses in Educational Research."

Background. Meta-analysis (Glass, 1976) is an increasingly popular statistical tool in many research fields (Schulze, 2007) to draw overall conclusions from different independent studies. Since conclusions of meta-analyses are often used in decision making

for policymakers and clinical practitioners (Borenstein, Hedges, Higgins, & Rothstein, 2009), it is important that meta-analytic conclusions are derived from the appropriate statistical models. Over the years, several new meta-analytic models have been developed, extending the range of research questions that can be answered. Examples of such techniques are network meta-analysis (Lumley, 2002), P-curve analysis (Simonsohn, Nelson, & Simmons, 2014), and meta-analytic structural equation modeling (Cheung & Chan, 2005; Jak, 2015; Viswesvaran & Ones, 1995). By advancing the meta-analytic techniques, the credibility and generalizability of meta-analytic conclusions may be increased (Ahn, Ames, & Myers, 2012).

In order to evaluate the performance of new extended meta-analytic techniques, researchers conduct simulation studies. In a simulation study, the meta-analytic technique under study is applied on several datasets that are randomly generated under some specific population model in different conditions of interest. For example, one may evaluate the performance of some model in conditions with different sample sizes, different numbers of effect sizes, different numbers of variables, and different population values. The results obtained in each generated dataset can consequently be compared to population values. To generalize the results of a simulation study, it is important that the data are generated under conditions that correspond to realistic research situations.

In educational research, there is insufficient information about typical meta-analytic conditions that can be used for simulation studies. Although there exist overviews of meta-analytic practices in educational research (Ahn et al., 2012; Polanin, Maynard, & Dell, 2017), these overviews mainly provide advice for future meta-analysists based on current practices. Both overviews do not report sufficient information to derive the typical meta-analytic conditions that can be used for simulation studies on meta-analytic techniques.

Objectives. Our research objective is to identify a comprehensive and recent overview of meta-analytic conditions in educational research that can be used for future simulation studies on meta-analytic techniques. Specifically, we focused on typical conditions to be used in simulation studies involving meta-analytic structural equation modeling.

Research question. The research question is: What are typical conditions of meta-analyses in educational research published between 2010 and 2017?

Method/Approach. We screened all 143 articles that were published in the journal 'Review of Educational Research' between March 2010 and February 2017. This journal is officially affiliated with the American Educational Research Association and has the highest impact factor (i.e., 5.6) for journals focusing on reviews in educational research (Web of Science, 2017). In total, 14 meta-analyses met the inclusion criteria (i.e., the article includes a meta-analysis, and the pooled effect size is a correlation coefficient). For each included meta-analysis, we coded the relevant characteristics, such as the number of variables of interest, the number of observed effect sizes, the sample sizes, the estimated pooled effect sizes, and whether the research questions involved mediation and/or moderation effects. Then, across the included meta-analyses, we calculated the mean, median, minimum, and maximum values of the relevant characteristics. We defined the 'typical' meta-analysis, as a meta-analysis with characteristics corresponding to the median

value across included meta-analyses.

Results/Findings. A typical meta-analysis includes 44 independent samples, with a total of 37150 participants. Of all primary studies included in a typical meta-analysis, the minimum sample size is 72, the median sample size is 422, the mean sample size is 1299, and the maximum sample size is 18687. A typical meta-analysis investigates a relation between three variables. The minimum pooled correlation coefficient in a typical meta-analysis is .16, the median and mean pooled correlation coefficient is .23, and the maximum pooled correlation coefficient is .33. All except one meta-analyses included a moderation analysis. More than half of the included meta-analyses are interested in testing relations between more than two variables, and approximately one-third of the included meta-analyses may (also) be interested in testing mediation.

Conclusions and Implications. This meta-meta-analysis presents an overview of typical meta-analytic conditions in educational research that can be used for future simulation studies on meta-analytic techniques. Moreover, our overview showed that there is often a mismatch between the research question and the analysis model. Most of the included meta-analyses arose from complex hypotheses, such as models that involve relations between a set of variables, and mediation models. Although these types of models are ideally evaluated using multivariate analysis, or meta-analytic structural equation modeling (MASEM), 13 of the 14 meta-analyses used univariate meta-analysis instead.

One reason why researchers used suboptimal univariate meta-analysis, may be that current MASEM-methods to evaluate the influence of continuous moderators are still limited (Cheung & Cheung, 2016). In our overview, most meta-analyses performed at least one moderation analysis. If the lack of possibilities to evaluate moderator effects in MASEM is indeed holding back researchers from using these techniques, it is essential that MASEM methods be further developed regarding possibilities to apply moderation analysis. Our current research will benefit such future developments by providing the typical meta-analytic conditions to be used in simulation studies.

References

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82*, 436-476. doi:10.3102/0034654312458162

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* Chichester, United Kingdom: John Wiley & Sons.

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological methods*, *10*, 40-64. doi:10.1037/1082-989X.10.1.40

Cheung, M. W. L., & Cheung, S. F. (2016). Random‐effects models for meta‐analytic structural equation modeling: Review, issues, and illustrations. *Research synthesis methods*, *7*, 140-155. doi:10.1002/jrsm.1166

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *The Educational Researcher, 10*, 3–8. doi:10.3102/0013189X005010003

Jak, S. (2015). *Meta-analytic structural equation modelling*. Springer International Publishing.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine, 21*, 2313-2324. doi:10.1002/sim.1201

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2016). Overviews in Education Research A Systematic Review and Analysis. *Review of Educational Research, 87*, 172-203. doi:10.3102/0034654316631117

Schulze, R. (2007). The state of the art of meta-analysis. *Journal of Psychology, 215*, 87-89. doi:10.1027/0044-3409.215.2.87

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology, 143(2),* 534-547. doi:10.1037/a0033242

Web of Science. (2017, November 12). Retrieved from https://webofknowledge.com

Viswesvaran, C., & Ones, D. (1995). Theory testing: Combining psychometric meta-analysis and structural equation modeling. *Personnel Psychology, 48*, 865–885. doi:10.1111/j.1744-6570.1995.tb01784.x

5 pm - 5:30 pm

Michael Schneider, Franzis Preckel

"Instruction- and student-related correlates of achievement in higher education: A systematic review of meta-analyses."

Background. In most industrialized countries across the world, close to 40% of the 25–34-year-old citizens have completed tertiary education. Persons with a degree in higher education tend to have better results in adult literacy tests, a lower chance of unemployment, and better health than their peers. At least partly, these are causal effects of education rather than mere correlates. A key question in the design of effective higher education concerns the sources of students' academic achievement. Which characteristics of students, teachers, and instruction are associated with higher learning outcomes than others? In our study, we use this definition of academic achievement: "performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university. […] Among the many criteria that indicate academic achievement, there are very general indicators such as procedural and declarative knowledge acquired in an educational system [and] more curricular-based criteria such as grades or performance on an educational achievement test" (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014). Previous reviews of meta-analyses on correlates of achievement focused on school learning or on education in general (e.g., Hattie, 2009), but higher education differs from K-12

school learning on a number of dimensions. Therefore, we conducted a review of all meta-analyses on variables associated with achievement in higher education.

Objective. Objective of the study was to compile and synthesize the published meta-analytic evidence on correlates of academic achievement in higher education.

Research questions. Our research questions were: (1) What environment- and learner-related correlates of achievement in higher education have been examined in meta-analyses so far? (2) What is the rank order of these correlations? (3) How many strong, medium, or weak effect sizes can be found for frequently investigated categories of environment- and learner-related variables, such as, e.g., instructional technology, teacher presentation, and student motivation?

Method. In April 2015, we systematically searched the titles, abstracts, and keywords of all articles in the literature database PsycINFO using the search string *(achievement or grades or competence or performance or learning or GPA) and ("higher education" or college or university or tertiary)* and limited the results to the meta-analyses published in English in peer-reviewed journals. In addition to the standardized search, we conducted an exploratory search on GoogleScholar and by scanning the reference lists of relevant reviews, books, and articles. To avoid overlap between the meta-analyses, we included only the largest meta-analysis on each topic, which was usually also the most recent one. Of the 124 found articles, 86 were excluded because they did not fulfill our inclusion criteria, i.e. were no meta-analysis, were not the largest meta-analysis on a topic, did not include an achievement measure, had less than 50% participants in higher education, or were limited to a single subject or subpopulation. Each co-author coded about half of the meta-analyses. Twenty effect sizes were coded by both co-authors independently, and the inter-coder agreement was 100%. To present all effect sizes in a common metric, we converted Pearson correlations to Cohen's *d*'s by using the formula . We classified each effect size as either indicating no effect ($|d| < 0.11$) or a small ($0.11 \leq |d| < 0.35$), medium ($0.35 \leq |d| < 0.66$), or large ($|d| \geq 0.66$) effect. Our cutoff values for these categories were based on Cohen's (1992) suggestion that effect sizes around $d = 0.20$ should be interpreted as small, those around $d = 0.50$ as medium, and those around $d = 0.80$ as large. We used the arithmetic means of neighboring values as category boundaries.

Results. The 38 included meta-analyses had been published between 1980 and 2014, with 23 meta-analyses published over the last 10 years (2005 to 2014). They investigated the correlations between 105 variables and achievement in higher education, based on a total of 3,330 effect sizes and involving an estimated total of 1,920,239 participants. Each meta-analysis reported effect sizes for one to three variables, except for the analyses of Poropat (2009; 6 variables), Robbins et al. (2004; 8 variables), Feldman (1989; 13 variables), and Richardson et al. (2012; 38 variables). In order to integrate the findings and to aid interpretation we identified each variable as instruction-related or student-related and assigned it to one of eleven categories, which correspond to different strands of the

research literature. Within the instruction-related and student-related variables, the categories were ordered by the combined frequency of medium and large effects. Among the instruction-related variables, social interaction and the stimulation of meaningful learning are the types of instructional practices most strongly associated with achievement. Assessment and presentation techniques were about equally important. Instructional technology and extracurricular trainings had the lowest effect sizes. Among the student-related variables, intelligence, prior knowledge, and learning strategies were most closely related to achievement. Personality variables and context variables are well investigated, but had comparably low effect sizes.

Conclusions and Implications. Due to space restriction, we can only give a superficial summary of the results here. As more detailed analyses in our study revealed, teachers with high-achieving students stimulate social interaction in the classroom, invest time and effort into planning and organizing their courses, provide clear learning goals and class objectives, give task-focused improvement-oriented feedback, and treat students with friendliness and respect. Instructional technology had comparably modest effect sizes, that did not increase over the past decades. Students in higher education with above-average achievement have high self-efficacy, intelligence, and prior achievement, miss few classes, engage in effort regulation, and use learning strategies in a goal-directed way. Most meta-analyses averaged over experimental and correlational studies, leaving it open whether causal relations underlie the reported correlations. However, overall, achievement in higher education is well researched and understood. Policy makers and teachers in higher education can improve the effectivity of classes by adhering to evidence-based instructional design principles.

References

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. *Oxford Bibliographies.* Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml

5:30 pm - 6 pm

Bianca Annabelle Simonsmeier

“Domain-Specific Prior Knowledge and Learning: A Meta-Analysis.”

Background. Domain-specific prior knowledge has been hypothesized to be one of the strongest positive determinants of learning. Previous studies have found that prior

knowledge can positively affect learning mediated through some mental processes (Dochy, Segers, & Buehl, 1999; Hambrick & Engle, 2002), but can negatively affect learning mediated through other mental processes (Luchins & Luchins, 1987; Vosniadou, Vamvakoussi, & Skopeliti, 2008). The mechanisms mediating the positive or negative effects of prior knowledge on learning do not mutually exclude each other. They might work in parallel and interact with each other. This interplay raises questions about the combined net effect of prior knowledge on learning. Is it positive, negative, or close to zero because positive and negative influences cancel each other out?
Objectives. Given the amount and heterogeneity of empirical findings, a meta-analytic integration of the empirical evidence would be helpful for a comprehensive understanding, but no such meta-analysis has been published so far. To close this gap we synthesized the published quantitative evidence on the relations between domain-specific prior knowledge and learning outcomes by means of a meta-analysis.

Research questions. The present meta-analysis followed four research questions. Our first research question related to how much empirical evidence is available about the relation between prior knowledge and both posttest knowledge and knowledge gains, and about the causality of these relations. As prior knowledge has widely been used throughout psychology and education for decades, we expected to find a large amount of relevant empirical evidence. Our second research question was about how prior knowledge relates to posttest knowledge. We hypothesized a strong positive correlation between prior knowledge and posttest knowledge because learners' knowledge in a domain typically accumulates over months or years. Our third research question concerned how prior knowledge predicts pretest-posttest knowledge gains. We expected to find a weak positive correlation between prior knowledge and knowledge gains because the amount of knowledge gained during a learning phase depends on a multitude of instruction characteristics, teaching characteristics, and learner characteristics (e.g., Ormrod, 2012). Our fourth research question related to how strongly the effects of prior knowledge on posttest knowledge and on knowledge gains are modulated by characteristics of the knowledge, the learner, and the environment.

Method. We performed a standardized search in the literature database PsycINFO in May 2015 and May 2017 which provided 5462 search results in total. After coding of abstracts and full-texts based on our inclusion criteria and coding rules, 240 studies were included in the meta-analysis. The 240 included articles reported results from 335 independent samples with 4327 relevant effect sizes and 62,129 participants in total.

We conducted the meta-analysis using robust variance estimation (Tanner-Smith, Tipton, & Polanin, 2016), which permits the inclusion of statistically dependent effect size estimates in a single meta-analysis without requiring information about the inter-correlation between effect sizes within studies. Given the presumed heterogeneity, random effects statistical models were used for all analyses. To address study artifacts that alter the value of outcome measures, we made corrections to the correlations obtained from

the single studies for measurement error and dichotomization (Schmidt & Hunter, 2015). We identified outliers through Cook's values (Cook & Weisberg, 1982; Viechtbauer & Cheung, 2010). We visually and statistically tested for publication bias (Duval & Tweedie, 2000; Egger, Smith, Schneider, & Minder, 1997).

Findings. Relating to our first research question, a large number of studies reported the correlation between prior knowledge and posttest knowledge. Of these, 44 effect sizes from nine studies were obtained in randomized controlled trials (RCTs). For 28 studies with 1305 effect sizes, it was possible to control for intelligence using partial correlations. In contrast, the association between prior knowledge and knowledge gains had been investigated in only a few studies, none of which was an RCT or allowed controlling for intelligence.

With respect to our second research question, we found a positive and statistically significant correlation with $r_P^+ = .525$. As expected, the correlation was strong according to the standards set by Cohen (1992), indicating a high stability of individual differences in knowledge from before to after learning. The connection between prior knowledge and posttest knowledge is causal, as indicated by the significant positive effect size found in the RCTs. Controlling the correlation for intelligence did not lead to a statistically significant decrease.

With respect to our third research question, we expected to find weak positive effects of prior knowledge on knowledge gains. On the group level, we found strong pretest-posttest knowledge gains of Cohen's $d = 1.62$. However, the empirical findings with respect to the individual differences were inconclusive. The correlations between prior knowledge and gain scores were descriptively negative, but not significantly different from zero. The 95% confidence intervals around these means were extremely large because only few studies reported the correlation between prior knowledge and knowledge gains, which makes it impossible to draw any definitive conclusions about how strongly prior knowledge correlates with knowledge gains. Due to a lack of relevant empirical studies, the causality of the prior knowledge effects on knowledge gains and the influence of intelligence could not be investigated in our meta-analysis.

Our fourth research question concerned how strongly the effect of prior knowledge on posttest knowledge and on knowledge gains is moderated by third variables. We found significant moderating effects for knowledge-related moderators (knowledge type and similarity of the prior knowledge measure and the posttest measure), learner –related moderators (participants 'educational level), and environmental-related moderators (cognitive demands of the interventions).

Conclusion and implications. Overall, the high stability of individual differences in knowledge supports theories emphasizing the accumulative long-term nature of knowledge acquisition. More randomized controlled trials investigating knowledge gains are needed. Generally, several processes (e.g., encoding and elaboration) mediate the effect of prior

knowledge on learning. The moderators of prior-knowledge effects need to be interpreted in terms of which mediating processes they affect. The discussion elaborates on these issues and offers a framework for further research on learning through knowledge acquisition.

6 pm - 6:30 pm

Peter Edelsbrunner, Christian M. Thurn

"The Prevalence of Unfounded (Statistical) Inferences Based on non-significant p-values in Educational Psychology."

Background. p-values as statistical tool for testing the predictions of hypotheses have been frequently criticized in recent years. Review studies and methodological discussions have shown up that researchers frequently make mistakes when interpreting p-values, particularly when they indicate a non-significant result. The more important question, we argue, is not how to interpret p-values, but to gauge the theoretical and practices consequences of misinterpretations.

Research questions. We review two misinterpretations of non-significant p-values in recent research (year 2016) on educational psychology, and try to gauge the consequences for theory and practical recommendations. The first misinterpretation is that non-significant p-values indicate the absence of an effect (Altman & Blend, 1995), and the second misinterpretation is that one effect with a significant p-value and another effect with a non-significant p-value indicate that the two effects differ from each other (Gelman & Stern, 2006). The research question is encompasses three target journals, with low- medium- and high impact factor (IF).

Method. We systematically reviewed all empirical articles from the 2016 volumes of the

German Journal of Educational Psychology (low 5-year IF), Instructional Science (medium 5-year IF), and Journal of Educational Psychology (high 5-year IF). For an initial equally balanced comparison of findings from the three journals, 10 articles were randomly drawn from each journal that are presented here (an overview of the findings from all articles will also be shown).

Results. We first present an overview of 30 randomly drawn articles, ten from each journal volume. Overall, out of 238 p-values across 30 articles there were 95 p-values > .05 (which was always the pre-specified cut-off for significance). Regarding the first misinterpretation, all these p-values > .05 were statistically misinterpreted as implying evidence for the absence of an effect. Regarding the second misinterpretation, 58 times researchers interpreted a difference between an effect with p > .05 and another one with p < .05 without conducting an appropriate test. The frequency and nature of misinterpretations differed between journals, the potential reasons for which will be discussed. Regarding gauging the theoretical and practical consequences, there was a great variety of implications that researchers inferred based on these misinterpretations. A common practice was that in the case of p > .05, researchers discussed reasons for the

absence of an effect. Regarding practical implications, the second misinterpretation beard more weight than the first one: For example, in various cases researchers pointed out the remarkableness of a significantly effective educational intervention training over another intervention that did not yield p-values below the cut-off. In some articles, authors raised the assumption that p > .05 might stem from small sample sizes. While sample size is linked to the frequency of a type II error, in none of the articles a potential type II error was explicitly discussed. In some articles, effects with p > .05 were neglected and not discussed.

Conclusions and implications. Misinterpretations of p-values commonly exist in the 2016 volumes of three journals with different IF from educational psychology. In addition, misinterpretations frequently lead to unfounded theoretical conclusions and suggestions for practical implications. We suggest more careful interpretations of p-values > .05, additional and alternative analyses strategies, and increased awareness.