

Advances in Investigating Response Behavior

Symposium title

Advances in Investigating Response Behavior

Coordinator

Marie-Ann Sengewald & Esther Ulitzsch

Affiliation

Leibniz Institute for Educational Trajectories (LifBi), Germany; University of Oslo, Norway

Abstract

Understanding how individuals engage with items in cognitive assessments and questionnaires holds immense potential to inform the design of assessments and ensure the validity of conclusions drawn from them. From a substantive point of view, modeling response behavior is also interesting in its own right. Such models tap on individual characteristics measured by behavioral indicators that could serve as theoretically relevant dependent or independent variables in substantive applications. This symposium presents six recent developments in modeling and investigating response behavior in both cognitive and non-cognitive assessments. The first three studies focus on careless and insufficient effort responding in questionnaires. Study I (Deribo & Kroehne) sheds light on its occurrence and investigates its relationship with the time of the day. Study II (Kroehne, Persic-Beck, Tetzlaff, Hahnel, & Goldhammer) explores its prevention by investigating whether real-time feedback following the detection of careless responding impacts subsequent response behavior. Study III (Uglanova, Nagy, & Ulitzsch) focuses on the detection of careless and insufficient effort responding, and evaluates the trustworthiness of model-based identification by drawing on experimental survey data. The second set of studies leverages process data from cognitive assessments to enhance psychometric and substantive research by offering deeper insights into examinees' response processes and skills. Study IV (Nagy & Ulitzsch) proposes a mixture item response theory model integrating response times to model and investigate partial disengagement as an extended classification of response behavior. Study V (Welling, Ulitzsch, & Nagy) integrates information on page revisits with item response theory models to explore the meaning of text re-reads in literacy assessments. Study VI (Sengewald, Torkildsen, Kristensen, & Ulitzsch) examines strategies for incorporating disengagement indicators into differential effect analyses, focusing on the benefit of process information for evaluating app-based training effects. The strategies and models presented in this symposium offer innovative perspectives on detecting and preventing careless and insufficient effort responding, as well as examining the value of response behavior analysis for advancing cognitive and non-cognitive assessments and substantive research.

Keywords

Response Behavior; Psychometric Analysis; Validation

Number of communications

6

Communication 1

Authors

Tobias Deribo & Ulf Kroehne

Affiliation

Leibniz Institute for Research and Information in Education (DIPF), Germany

Abstract

Questionnaires are a cornerstone of scientific research when wanting to measure non-cognitive constructs. However, low motivation to complete them can lead to improper responses and compromise the validity of the drawn conclusions (i.e., Maniaci & Rogge, 2014; Podsakoff et al., 2012), especially when using unsupervised online formats (i.e., Kroehne et al., 2020). One specific factor related to response motivation may be the time of day the questionnaire was undertaken (i.e., Kouchaki & Smith, 2014; Olsen et al., 2017). Here, prior research specifically points to night-time as a risk factor, as it may be coupled with increased exhaustion, tiredness, and depleted cognitive resources (i.e., Dickinson and Whitehead, 2015). Furthermore, this effect may be enhanced when surveys are work-related but completed outside the professional context. Therefore, this study wants to investigate how the time of survey processing impacts unmotivated response behavior in the form of Careless and Insufficient Effort Responding (C/IER; Huang et al., 2015) and if there are other variables related to the time of survey processing.

The analysis draws on data from $N = 2,697$ teachers participating in an online questionnaire to measure school and teaching development conditions. Teachers provide an interesting sample here, as they are often susceptible to work outside of regular working hours (e.g., Forsa, 2022). Various types of unmotivated response behavior, such as straight-lining or rapid responding, were identified as indicators of low response motivation (Curran, 2016). From these indicators, we derived variables for the appearance and the number of unmotivated responses per respondent.

To address the research question, a Bayesian zero-inflated negative binomial regression (Loeys et al., 2012) was applied to predict the appearance and the number of unmotivated responses. Predictors were gender, measures for job exhaustion and job satisfaction, as well as time of survey processing. The model was estimated with weakly informative priors and four chains with 5,000 iterations (half as burn-in). The applied model converged well with a Potential Scale Reduction Factor of < 1.1 for all modeled parameters and an Effective Sample Size $> 1,000$ (Bürkner, 2017). The Region of Practical Equivalence criterion (Kruschke, 2018) was applied to judge the practical significance of the obtained results.

The results reveal that surveys completed outside regular working hours, or the workweek are associated with a higher likelihood of C/IER. However, no practically significant effects were found on the absolute number of unmotivated responses. These results provide insights into potentially undesirable consequences of unsupervised online surveys. Furthermore, they indicate that enabling participants to complete work-related surveys during regular working hours appears advisable to draw a higher number of valid conclusions from survey data.

This study underscores the role of time of day as a helpful design aspect for preserving data quality and minimizing the risks posed by C/IER. By addressing aspects of participant motivation as early as the design phase, researchers can create conditions that support higher data quality.

Keywords

Time of Day; Online Research

Communication 2

Real-time detection of unmotivated response behavior in questionnaires - Can immediate feedback influence future response behavior?

Authors

Ulf Kroehne, Lothar Persic-Beck, Leonard Tetzlaff, Carolin Hahnel & Frank Goldhammer

Affiliation

Leibniz Institute for Research and Information in Education (DIPF), Germany

Abstract

Unmotivated responses, identified using response times (as rapid guessing in cognitive tests, Wise & Kong, 2005; or as rapid responding in questionnaires, as part of the careless and insufficient effort responding, C/IER), are a known threat to validity (e.g. Wise, 2017). It is known from the literature that unmotivated response behavior occurs more frequently in low-stakes assessments (Wise et al., 2009), for male test takers or respondents (e.g. DeMars, Bashkov, & Socha, 2013) and with increasing item numbers (Lindner et al., 2019). However, specific psychometric models for identified rapid responses at item level (e.g., Deribo et al., 2021) or incorporating response time effort (RTE) as a process indicator at person level can only indirectly improve data quality and the validity of measurements as a post hoc correction based on already contaminated data. This paper examines how real-time detection of unmotivated response behavior during the data collection is possible and whether immediate feedback on the observed unmotivated response behavior as micro-intervention influences future responding. Using an experimental design with between-subject variation, feedback on leaving a questionnaire page that indicates missing answers, monotonous (i.e., “straightlining”) or rapid (i.e., “rapid responding”) answers in a computerized questionnaire (experimental group) is compared with feedback that only indicates missing answers (control group). The questionnaire providing log event data necessary for the identification of item-level response times in questionnaires with several items per page was administered in a nation add-on study to PISA 2022 (N=705). The position of contiguous questionnaire screens, each containing one scale, was counterbalanced using a balanced design with 18 booklets. Real-time detection of rapid responding was implemented using algorithmic processing of log events (Kroehne & Goldhammer, 2018) to extract the average answering time (AAT). AAT is known from previous analyses (Kroehne et al., in press) to show a bimodal distribution in the presence of rapid response behavior, and a conservative time threshold of 1.0 seconds was chosen for the detection of rapid response behavior in the experimental condition. The results confirm the expected bimodal distribution of the AAT, supporting the two hypothesized response processes in the experimental condition and control group. For both male and female 15-year-old students, a significant effect of the feedback on the average response time and the probability of showing quick response behavior (standardized odds ratio of 0.867 for boys and 0.734 for girls) was found. In addition to the direct effects of the micro-intervention, which remains significant when controlled for position effects, we report further indirect effects on data quality (reliability, differential item functioning and latent correlations), and present descriptive results of an inserted in-situ question to test takers on how the identified responses should be used. While the real-time detection of unmotivated response behavior affects future response behavior, the overall effect sizes of the micro-intervention are low. In the concluding section, the practical significance of the results for future computerized surveys is discussed.

Keywords

Real-Time Detection; Logdata; Computerized Surveys

Communication 3

Experimental Validation of Model-Based Identification of Careless and Insufficient Effort Responding

Authors

Irina Uglanova, Gabriel Nagy, & Esther Ulitzsch

Affiliation

Leibniz Institute for Science and Mathematics Education (IPN), Germany

Abstract

Self-report surveys often suffer from careless and insufficient effort responding (C/IER), which refers to responses provided without paying attention to the items' content. Mixture modeling approaches are promising

tools to assess C/IER by means of latent class variables. However, evidence for the validity of interpreting the latent class variable as C/IER is still pending. To shed more light on this issue, this paper presents the results of a pre-registered survey experiment. Specifically, we examine the ability of a recently developed mixture item response theory model (Uglaova, Nagy, & Ulitzsch, in preparation) to detect C/IER in self-report measures of dark personality traits. We evaluated two validity arguments: the relationships of the latent class (1) with experimentally manipulated survey conditions and (2) with alternative indicators of C/IER. Experimental conditions were designed to evoke or prevent C/IER by manipulating the instructions, the presence of cognitively demanding tasks, and the number of preceding items. Alternative indicators of C/IER were attention check items, item content recognition tasks, and self-reported C/IER. Using the bias-adjusted three-step approach to relate latent class membership to external variables we found that (a) respondents in the evoking C/IER condition were more likely to be assigned to the C/IER class than respondents in the preventing condition, and (b) respondents assigned to the C/IER class exhibited lower performance on all alternative indicators of C/IER than those in the attentive class. Overall, these results were consistent with our pre-registered hypotheses, providing validity arguments to support interpreting the latent class variable as representing C/IER.

Keywords

Experimental Validation; Mixture IRT Model

Communication 4

A Multilevel Mixture Item Response Theory Model for Partial Engagement in Proficiency Tests

Authors

Gabriel Nagy & Esther Ulitzsch

Affiliation

Leibniz Institute for Science and Mathematics Education (IPN), Germany

Abstract

Disengaged test-taking behavior is a problem in low-stakes assessments. To account for low engagement, popular approaches rely on item response times to classify responses as disengaged (rapid guessing) or inconspicuous (engaged). Although conceptually elegant, this binary classification has been found to miss a substantial proportion of disengaged responses. This paper introduces an extended classification of engagement that includes “partial (dis)engagement”. To this end, a Multilevel Mixture Item Response Theory (MMIRT) model is proposed that classifies engagement at the item level. Partially engaged responses are specified to be associated with response times that fall between the very short response times of rapid guesses and the response times of engaged responses. Responses are classified on the basis of within-individual response time distributions, meaning that the model accounts for individual differences in habitual time expenditure. Disengaged responses are modeled as the result of a guessing process, while partially and fully engaged responses are both related to the proficiency variable via a three-parameter response model. The MMIRT model can be estimated using maximum likelihood techniques via the expectation maximization algorithm. The MMIRT model is illustrated with data from the TIMSS 2019 science assessment. Multiple-choice items presented at the beginning and end of the test in a rotated test design were analyzed. In the U.S. sample of eighth graders, test performance was lower on items presented at the end of the test. The lower performance could not be explained by models based on the binary classification of response engagement. In contrast, the proposed MMIRT model suggested that the decline in performance was due to an increase in partially disengaged responses.

Keywords

Multilevel Mixture IRT; Engagement Classification

Communication 5

Investigating the interplay of text rereads with IRT parameters: Rereads render hard items easier and easy items harder

Authors

Jana Welling, Esther Ulitzsch, & Gabriel Nagy

Affiliation

Leibniz Institute for Educational Trajectories (LifBi), Germany

Abstract

In reading comprehension tests, test-takers can choose to reread the text of the task while working on an item. Up to now it is not well understood how rereading the text relates to test performance and its measurement. To close this gap, the aim of the present study was to investigate the relationship between text rereads on one hand and item parameters of item response models and test performance on the other hand. We specified three different item response mixture models that distinguish on the response level between the three latent classes rapid guessing, solution behavior with text rereads and solution behavior without text rereads. The different models assumed either (1) equal item parameters, (2) equal item discriminations but varying item difficulties, or (3) varying item parameters between the two different solution behavior classes. In a reading comprehension test of the German National Educational Panel Study (N = 1933 students, 14 multiple-choice items), the second model with equal item discriminations but varying item difficulties between the two latent classes fitted the data best. Descriptive analysis revealed that the reread class did not differ extensively from the no reread class in the average item difficulty, but rather exhibited less variation in the item difficulties, rendering hard items easier and easy items harder. Furthermore, the tendency to reread the text positively predicted test performance. The results highlight the importance of investigating process data beyond item response times, which can help to better understand the test-taking process as well as its interplay with the measurement of test performance.

Keywords

Text Rereads; Mixture IRT Model

Communication 6

Benefits of Process Data for Evaluating the Differential Effectiveness of App-Based Treatments

Authors

Marie-Ann Sengewald, Janne v. K. Torkildsen, Jarl K. Kristensen, & Esther Ulitzsch

Affiliation

Leibniz Institute for Educational Trajectories (LifBi), Germany

Abstract

Differential effect analysis can reveal the preconditions for effective interventions by highlighting variations in intervention outcomes. The growing use of digital tools, such as learning apps, provides rich process data on response times and response behavior, offering insights into how participants interact with these apps. We use this information source and bridge psychometric research on controlling for disengaged responding with differential effect analysis to evaluate how variations in the usage of learning apps contribute to heterogeneity in intervention effectiveness. Specifically, we consider different response-time-based indicators to identify disengaged behavior, including thresholds for overly short response times, Gaussian mixture modeling, and model-based approaches that integrate item responses and response times (e.g., Wise & Kong, 2005; Wise, 2017; Ulitzsch et al., 2020; 2023). We demonstrate how these indicators can be integrated into the EffectLiteR

framework, which specifies a structural equation model for differential effect analyses with latent variables (Mayer et al., 2016; Sengewald & Mayer, 2024). Finally, we compare the different modeling strategies and investigate the benefits of using the disengagement indicators in an empirical application. For this, we rely on the work of Torkildsen et al. (2022), who constructed an app-based morphological training program and evaluated its effectiveness in a randomized controlled trial with 717 second-grade students. Using the empirical data, we examine the heterogeneity of the morphological training effects in relation to the pre-treatment characteristics of the students and the gains achieved by including the different disengagement indicators, focusing on their impact on explained outcome variance and effect size differences. Our findings identify baseline characteristics that predict greater benefits from the training and highlight how different modeling strategies for disengagement indicators influence the conclusions. Beyond the practical insights into the utility of process data, the results demonstrate the application of the advanced modeling strategies for differential effect analysis.

Keywords

Differential Effectiveness; Disengagement Indicators; EffectLiteR

Primary author: SENGEWALD, Marie-Ann (Leibniz Institut for Educational Trajectories)

Co-author: ULITZSCH, Esther (University of Oslo)

Presenter: SENGEWALD, Marie-Ann (Leibniz Institut for Educational Trajectories)

Session Classification: Symposium : "Advances in Investigating Response Behavior"

Track Classification: Measurement: Measurement