

Text Analytic Techniques in Survey Questionnaire Development and Analysis

John Ford

U.S. Merit Systems Protection Board

Abstract: This research develops three text analytic techniques to improve survey questionnaires. The first is open-ended response mining. Narrative responses on a survey are mined for themes then used to develop new questions. Closed-ended responses identify subgroups who agree/disagree with the question. Then open-ended responses examined for systematic differences which suggest new constructs that distinguish the groups. The second is used during question development. Agree/disagree questions are examined for similarity in language using latent semantic analysis. The matrix of similarity coefficients is used to make scale assembly and predicted item performance decisions in advance of field test data. The third involves replacing zero with LSA-derived coefficients as baseline comparisons for correlation coefficients to identify interesting relationships between rating questions. Semantic similarity of question stems suggests a degree of relationship between questions. This, rather than zero, is the appropriate expected value of a correlation between two items.