Psychometric Evaluation of the 10-item Questionnaire of Smoking Urges – Brief (QSU-Brief)

Stacey McCaffrey, PhD, Andrea Vansickel, PhD, Ryan Black, PhD

Altria Client Services LLC, 601 East Jackson Street, Richmond, VA 23219, USA

SRNT 25th Annual Meeting, February 20 - 23, 2019, San Francisco, California, USA

Altria
Altria Client Services

This poster may be accessed at www.altria.com/ALCS-Science

Abstract

Significance: There are many brief versions of the Questionnaire of Smoking Urges (QSU) with different items, rating scales, and scoring. A commonly used version is the 2-factor 10-item "QSU-Brief". Despite its popularity, little is known about its psychometric properties and previous research has yielded inconsistent findings. Therefore, the purpose of this research is to generate guidance for researchers through a psychometric evaluation of the QSU-Brief (7-point rating scale (RS)).

Method: Using a Rasch modeling framework, we evaluated the QSU-Brief using data from 3 clinical studies (N=314) where adult cigarette smokers were randomized to their own brand cigarettes, e-vapor, or nicotine replacement therapy (NRT) gum.

Results: A principal components analysis on the probability scale residuals provided evidence of dimensionality. The first dimension represented "intention and desire to smoke" (IDS), while items 4, 8, and 9 loaded onto a second dimension, "relief of negative affect with an urgent desire to smoke" (RNA). IDS: We sequentially removed items 10, 5, and 2 due to problems with RS functioning and item fit. The remaining 4 items exhibited appropriate RS functioning, good reliability (person reliability =.86, separation=2.52), unidimensionality, and did not exhibit bias (evaluated via differential item function (DIF)) based on gender, race, or age. The person-to-item map suggested a ceiling effect. IDS differed between products. Change in IDS scores following ad lib use differed by product, whereby subjects randomized to cigarettes reported greater reduction in IDS than e-vapor or gum. Exploratory analyses suggested that change in IDS was related to amount of product used during ad lib use. RNA: We collapsed the 7-point RS to a 5-point RS due to disordered thresholds. Person reliability and separation were .76 and 1.79, respectively. The items did not exhibit bias. Results suggested a floor effect, limited person-to-item targeting, and provided some evidence of known-groups validity and responsiveness.

Conclusion: This is the first study to comprehensively evaluate this QSU-Brief. While results provide support for the 4-item IDS scale, psychometric properties of the 3-item RNA scale were generally less robust. Therefore, strengths and weaknesses of this scale should be carefully considered before use. Overall, findings presented here are intended to aid researchers in selecting appropriate metrics for future studies.

Methods

ANALYSES

Using classical and modern test theory approaches, we evaluated:

(1) dimensionality
(2) rating scale functioning
(3) information
(4) item coverage (person-to-item targeting)
(5) reliability (person reliability and separation)
(6) differential item function (DIF)
(7) known-groups validity
(8) responsiveness

Analyses were conducted in Winsteps version 4.0.0 (Linacre, 2017) and SPSS version 25 (IBM, 2017).

DATA SOURCE

3 clinical studies (N=314) where adult cigarette smokers were randomized to their own brand cigarettes (OBC), e-vapor, or nicotine replacement therapy gum (NRT gum).

INSTRUMENT

The QSU-Brief being evaluated in this study is the 10-item QSU-Brief developed by Cox and colleagues (Cox et al., 2001) utilizing the original QSU 1 (strongly disagree) to 7 (strongly agree) rating scale (Tiffany & Drobes, 1991). The items are listed in Table 1.

TABLE 1. QSU-Brief Item Content

Item	Item content
1	I have a desire to smoke right now
2	Nothing would be better than smoking right now
3	If it were possible, I probably would smoke right now
4	I could control things better right now if I could smoke
5	All I want right now is a cigarette
6	I have an urge for a cigarette
7	A cigarette would taste good now
8	I would do almost anything for a cigarette now
9	Smoking would make me less depressed
10	I am going to smoke as soon as possible

Rating scale: 1 (strongly disagree) to 7 (strongly agree)
Instructions: For each item, please indicate how you feel RIGHT NOW

Results

DIMENSIONALITY

A principal components analysis on the probability scale residuals (Bond & Fox, 2007) in Rasch-modeling software (Winsteps; Linacre, 2017) provided evidence of dimensionality.

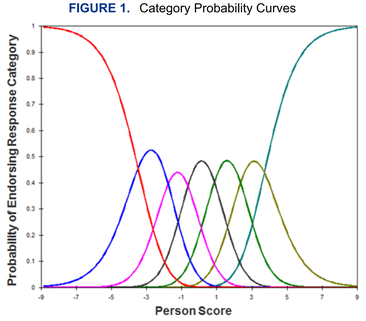
- The first dimension represented "intention and desire to smoke" (IDS), while items 4, 8, and 9 loaded onto a second dimension, "relief of negative affect with an urgent desire to smoke" (RNA).
- Results for these two dimensions are presented separately below.

Intention and Desire to Smoke (IDS)

RATING SCALE FUNCTIONING

We removed items 10, 5, and 2 due to problems with rating scale functioning and item fit. The remaining items (items 1, 3, 6, 7) exhibited appropriate rating scale functioning (Figure 1) and item fit.

FIGURE 1. Category Probability Curves



INFORMATION

Highest levels of information approximately 3 SD above and below the mean.

ITEM COVERAGE

Good person-to-item targeting and a potential ceiling effect.

RELIABILITY

Good reliability (person reliability =.86, separation =2.52) (Linacre, 2018).

DIF

No substantial DIF based on gender, age (legal age to 30 years vs. >30 years) or race (White vs. non-White).

KNOWN-GROUPS VALIDITY

The OBC group reported significantly lower IDS scores following 4-hour ad lib product use than the e-vapor and NRT gum groups. The e-vapor group also reported significantly lower IDS scores than the NRT gum group.

RESPONSIVENESS

The OBC group exhibited a greater reduction in IDS than the e-vapor and NRT gum groups following 4-hour ad lib product use. We did not find positive linear relationships between IDS change scores (change in IDS scores following 4-hour ad lib product use) and amount of product used as anticipated.

Relief of Negative Affect with an Urgent Desire to Smoke (RNA)

RATING SCALE FUNCTIONING

To correct problems with rating scale functioning (Figure 2), we collapsed the 7-point rating scale to a 5-point scale (Figure 3) by collapsing categories 2-3 and 5-6.

FIGURE 2. Original Category Probability Curves

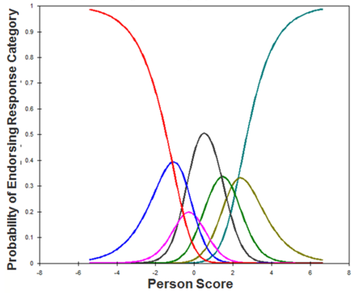
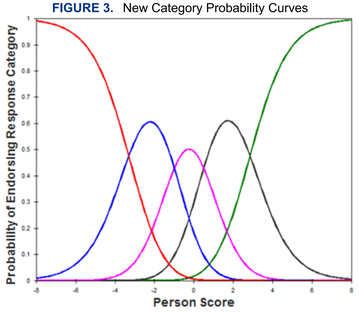


FIGURE 3. New Category Probability Curves



INFORMATION

Highest levels of information approximately 2 SD above and below the mean.

ITEM COVERAGE

Limited person-to-item targeting and a floor effect.

RELIABILITY

Reliability estimates were lower than expected (person reliability =.76, separation =1.79) (Linacre, 2018).

DIF

No substantial DIF based on gender, age (legal age to 30 years vs. >30 years) or race (White vs. non-White).

KNOWN-GROUPS VALIDITY

The OBC group reported significantly lower RNA scores following 4-hour ad lib use than the e-vapor and NRT gum groups. However, RNA scores did not differ between the e-vapor and NRT gum groups.

RESPONSIVENESS

The OBC group exhibited greater reduction in RNA than the e-vapor and NRT gum groups following product use. We did not find positive linear relationships between RNA change scores and amount of product used as anticipated, with the exception of time that subjects kept NRT gum in their mouths (r=.353, p=.047).

Discussion & Limitations

- This is the first comprehensive psychometric evaluation of the 10-item QSU-Brief proposed by Cox and colleagues (2001). Our analyses supported the removal of 3 poorly functioning items to arrive at a final 4-item IDS scale and 3-item RNA scale. Examination of RNA rating scale functioning revealed disordered Andrich thresholds, suggesting that participants were having difficulty discriminating between the response categories or were interpreting the response categories in meaningfully different ways. To achieve appropriate rating scale functioning, we collapsed the RNA rating scale from a 7-point scale to a 5-point scale during scoring. As this study did not incorporate qualitative research with consumers, we are limited in our ability to understand why several items and the original 7-point RNA rating scale were functioning poorly.
- This study relied exclusively on analysis of previous collected data, and we were not able to estimate test-retest reliability of the IDS and RNA scales with this data.
- IDS and RNA change scores were not related to product use variables as anticipated. Although exploratory analyses suggested possible nonlinear relationships, additional research with larger sample sizes may be needed to understand these findings.
- We evaluated known-groups validity and responsiveness within the context of 4-hour ad lib use of OBC, e-vapor, and NRT gum. Future research might evaluate known-groups validity and responsiveness within the context of different (1) time periods of exposure and (2) tobacco products.

Conclusion

Given the problems identified with the full 10-item QSU-Brief, these new proposed scales with Rasch-derived scoring may be considered as an alternative in future research.