*ets research institute

SEPTEMBER 2026
RM–26-07

RESEARCH MEMORANDUM

Addressing User Experience Concerns of Forced Choice: Introducing a Hybrid Multidimensional Forced-Choice/Likert (HMFCL) Response Format

AUTHORS

Kevin M. Williams, Diego Figueiras, and Katrina Roohr

ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey
Lord Chair in Measurement and Statistics

ASSOCIATE EDITORS

Usama Ali
Principal Measurement Scientist

Beata Beigman Klebanov
Principal Research Scientist, Edusoft/ETS

Katherine Castellano
Managing Principal Research Scientist

Michael Fauss
Research Scientist

Yang Jiang
Research Scientist

Jamie N. Mikeska
Managing Principal Research Scientist

Teresa Ober
Research Scientist

Jonathan Schmidgall
Senior Research Scientist

Jesse Sparks
Director Research

Zuowei Wang
Senior Research Scientist

Mikyung Wolf
Principal Research Scientist

Diego Zapata Rivera
Distinguished Presidential Appointee

Klaus Zechner
Senior Research Scientist

Jiyun Zu
Senior Measurement Scientist

PRODUCTION EDITORS

Ayleen Gontz
Manager, Editorial Services

Emily Ramsower
Manager, Editorial Services

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Addressing User Experience Concerns of Forced Choice: Introducing a Hybrid Multidimensional Forced-Choice/Likert (HMFCL) Response Format

Kevin M. Williams, Diego Figueiras, & Katrina Roohr

ETS Research Institute, ETS, Princeton, New Jersey, United States

September 2026

Corresponding author: Kevin M. Williams, E-mail: kmwilliams@ets.org

Suggested citation: Williams, K. M., Figueiras, D., & Roohr, K. (2026). *Addressing user experience concerns of forced choice: Introducing a hybrid multidimensional forced-choice/Likert (HMFCL) response format* (Research Memorandum No. RM–26-07). ETS. https://doi.org/10.64634/5r6xkc58

Find other ETS-published reports by searching the
ETS ReSEARCHER database.

To obtain a copy of an ETS research report, please visit
https://www.ets.org/contact/additional/research.html

Action Editor: Teresa Ober

Reviewers: Marisol Kevelson and Patrick Kyllonen

Copyright © 2026 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and FUTURENAV are registered trademarks of Educational Testing Service (ETS). All other trademarks are the property of their respective owners.

Abstract

The multidimensional forced-choice (MFC) response format for psychological testing was created to address social desirability and easy-to-fake concerns with other response formats such as the Likert-type response format. Although research suggests that MFC has largely achieved this goal, other issues such as negative user feedback hinder implementation efforts and jeopardize the authenticity of MFC data. We developed and tested a hybrid MFC/Likert (HMFCL) response format designed to provide a better user experience through greater response flexibility while maintaining the fake-resistance of the MFC format. This study describes user experience feedback from 53 participants in a within-subjects (i.e., A/B testing) design comparing the MFC and HMFCL formats. Quantitative and qualitative results suggest that respondents preferred the HMFCL format over traditional MFC on almost every metric (e.g., ease of use, fairness). Future directions for evaluating, implementing, and scoring data from the HMFCL format are discussed.

Keywords: HMFCL, forced-choice response format, Likert, personality, skills-based assessment, user experience, Futurenav™ Edge behavioral assessment

Introduction

The Likert response format (Table 1) is arguably the most utilized quantitative design in psychological assessment. Although there are some variations, this format generally requires individuals to rate their agreement with each of a set of statements on an ordinal scale (e.g., on a 5-point scale where 1 = *strongly disagree* and 5 = *strongly agree*). The Likert format is responsible for countless volumes of valuable psychological research knowledge over the past several decades. However, researchers have also noted that this format is susceptible to inaccurate responding in which test takers may rate themselves more highly on statements that are socially desirable, a phenomenon also known as *faking good* (see Williams, 2022). This type of responding is problematic because it introduces *noise* (i.e., confounding factors irrelevant to the target construct).

Table 1. Example Likert (Panel A) Items

Rate how much each of the following statements is like you or unlike you:				
Statement	Very Unlike Me	Unlike Me	Like Me	Very Like Me
I work well with others.				
I have a strong work ethic.				

Note. Traditional Likert format places no restrictions on participants' responses, and typically presents items in nonpaired format (e.g., single statement per page; lists of statements per page).

Though faking good may be influenced by dispositional characteristics (e.g., narcissism), practitioners in workplace settings are generally more concerned about temporary situational influences. For example, when completing assessments for workplace hiring purposes, individuals may misrepresent themselves to improve their chances of being selected (Hayes, 2007). These concerns jeopardize practical applications in which employers seek to obtain critical knowledge about applicants, such as personality or other noncognitive constructs or transferrable skills. These issues may disrupt employers' ability to hire and promote the most appropriate individuals. These concerns may extend to other workplace applications involving incumbent workers (e.g., performance reviews; see Williams, 2022) or beyond the workplace (e.g., college admissions; Krammer, 2020).

To address this concern, researchers have developed alternative response formats. One notable strategy is the forced-choice format (Table 2; Brown & Maydeu-Olivares, 2011; Fu, Kyllonen, & Tan., 2024). A common example of this format involves pairing two statements of similar social desirability with respondents selecting which statement better represents themselves (i.e., "best describes" them; Table 2). Because the two statements within a pair typically represent distinct constructs, this format is often described as multidimensional forced choice (MFC).

Table 2. Example Multidimensional Forced-Choice (MFC; Panel B) Items

Select which of the following two statements best describes you:		
I work well with others.	OR	I have a strong work ethic.

The rationale behind this format is that it decreases respondents' ability to simply respond in a socially desirable manner to all statements (Christiansen et al., 2005). For example, an item pairing may require a respondent to select between a statement that reflects high

teamwork skills and another that reflects high work ethic; the MFC format prohibits respondents from claiming that they demonstrate both of these traits. Research suggests that the MFC format has been largely successful in reducing faking concerns (e.g., Cao & Drasgow, 2019). Specifically, response formats' respective resistance to misrepresentation is often evaluated through faking studies, in which assessment scores are compared between individuals who are instructed to fake good (e.g., "Respond as though you are applying for a job") versus those who respond under standard, low-stakes instructions. These studies have demonstrated that MFC formats are less susceptible to faking (i.e., score inflation) relative to Likert (e.g., Christiansen et al., 2005; Jackson et al., 2000; see also meta-analysis by Cao & Drasgow, 2019).

We have utilized the MFC format in assessments such as the ETS Personal Skills and Qualities (PSQ) questionnaire and WorkFORCE Assessment for Job Fit, which demonstrates positive psychometric properties (Naemi et al., 2014). However, we have received qualitative feedback from test takers that is highly critical of the MFC format. Specifically, respondents often protest being forced to choose one statement within a pair when they feel that both statements are equally like them or unlike them. This feedback echoes those of previous studies, in which participants react negatively to the MFC format based on its restrictiveness and perceived inability to provide an accurate reflection of the test taker's personality (e.g., Borman et al., 2024; Dalal et al., 2021). Participants are particularly opposed to being required to endorse statements that may be worded in a socially undesirable fashion (e.g., "I'm often late for work"; "I don't like working with other people"). However, negatively worded statements such as these are argued as necessary to capture the entire possible spectrum of target constructs (e.g., personality). This argument is supported by research that demonstrates that samples generate higher variability in scores when negatively phrased items are included, particularly when antonyms are employed (as opposed to negations; Dueber et al., 2022; İlhan et al., 2024).

Regardless, this participant feedback is significant because it suggests a negative user experience, which jeopardizes adoption for both practical and research applications. Moreover, this negative feedback may adversely affect measurement accuracy. For example, when

respondents feel that both statements are equally like them or unlike them, they may simply select one statement at random. More generally, the negative user experience may promote disengagement among test takers, which may also lead to pattern responding or random responding (Cao & Drasgow, 2019). Cumulatively, these response patterns threaten measurement accuracy and validity.

To address this feedback, we developed a hybrid MFC/Likert format (HMFCL; Table 3). The rationale behind this new format is to maintain the forced-choice nature of the traditional MFC format while allowing respondents to communicate that there may not be a strict black-and-white distinction between the paired statements. This format presents paired statements, each of which is rated using a Likert-type response scale. Respondents are required to select a different response for each statement. For example, if a respondent selects Very Like Me for one statement, they cannot select this same response for the other statement. Thus, respondents are still forced to select which of the two statements is more like them, and this maintains the forced-choice nature of the traditional MFC format. However, this format also allows respondents to state that both statements are like them or unlike them to varying degrees, a feature that is absent in traditional MFC and directly addresses user criticisms (e.g., Borman et al., 2024; Dalal et al., 2021).

Table 3. Example Hybrid Multidimensional Forced-Choice Likert (HMFCL; Panel C) Items

Rate how much each of the following two statements is like or unlike you. *You cannot select the same rating for each statement:*				
Statement	Very Unlike Me	Unlike Me	Like Me	Very Like Me
I work well with others.				
I have a strong work ethic.				

Notably, HMFCL differs from response formats such as the graded forced-choice scale (Zhang et al., 2024), also referred to as the graded-pairs, graded-paired, or graded-preference format (Bech et al., 2007; Brown & Maydeu-Olivares, 2018; Lingel et al., 2024; Schulte et al., 2024). The HMFCL format measures the relative preference between two statements by requiring respondents to select which one is more like them (Tables 4 and 5). Thus, graded pairs typically assess the relative differences between two constructs. However, unlike HMFCL, graded pairs do not capture respondents' absolute level of each construct. Tables 4 and 5

depict an example of how the graded pairs format may fail to distinguish between distinct response patterns to the same extent as HMFCL. In this example, two respondents provide significantly different responses to the items as depicted in HMFCL: Respondent 1 selects Very Like Me and Like Me to statements A and B respectively, whereas Respondent 2 selects Unlike Me and Very Unlike Me to these same statements. That is, the two sets of responses represent opposite valences on the construct being measured. However, because the respondents demonstrate an identical relative preference to the two statements, their responses in a graded format are indistinguishable. Together, the examples in Tables 4 and 5 illustrate the additional critical information that may be captured through HMFCL that is absent in graded pairs. Moreover, the graded pairs format often includes a central response option (e.g., About the Same). This option not only negates the forced-choice nature of the format, but prevents any determination as to whether respondents select this option because they perceive both statements as like them or unlike them. Note: The examples in Tables 4 and 5 demonstrate how distinct HMFCL response patterns (Respondent 1 and Respondent 2) may not be detected in the graded pairs format.

Table 4. Hybrid Multidimensional Forced-Choice Likert (HMFCL) Response Format

		HMFCL			
Respondent	Statement	Very Unlike Me	Unlike Me	Like Me	Very Like Me
1	A				X
	B			X	
2	A		X		
	B	X			

Table 5. Graded Pairs Response Formats

		Graded Pairs					
Respondent	Statement	Much More	Slightly More	About the Same	Slightly More	Much More	Statement
1	A		X				B
2	A		X				B

There may be concerns that individuals may try to select the same response to each statement in the HMFCL format, possibly because they do not understand the restrictions required for HMFCL or may wish to disregard them. In digital data collection, the requirement that respondents select a different option for each response in HMFCL may be constrained

through the platform interface. For example, once a test taker has selected a response for the first statement in a pair, the interface may deactivate this same response option for the second statement. Alternatively, the interface may initially allow test takers to select the same response for each statement but may display an error message that prohibits them from advancing to the next pairing until they correctly select different responses for each statement.

In this paper, we describe a user experience study designed to formally evaluate users' perceptions of the HMFCL format relative to the traditional MFC format. This research was conducted as part of the early phases of product development for the Futurenav™ Edge behavioral assessment: a tool that utilizes the HMFCL and other response formats to measure employees' unique durable skills through behavioral and communication assessments. The data from this study are a first step in implementing the HMFCL format on a larger scale so that its other psychometric properties may be evaluated more thoroughly. We also examined timing data for each item in each response format, as these data will also impact practical applications and adoption.

Our primary research questions (RQs) for this study are as follows:

- RQ1: Do respondents report a more favorable user experience for the HMFCL format relative to traditional MFC?
- RQ2: How do administration times compare between HMFCL and MFC?

Methodology

Measures

For this study, we created two abridged versions of the Futurenav Edge behavioral assessment. The behavioral assessment was adapted from the foundational research from the ETS PSQ questionnaire (Naemi et al., 2014). Like the PSQ, the Futurenav Edge behavioral assessment is grounded in the five-factor model personality framework (i.e., agreeableness, conscientiousness, emotional stability, extraversion, openness to experience; see Naemi et al., 2014). Within the assessment, the naming conventions for the five factors were converted into competencies more appropriate for the workplace context: teamwork, diligence, stress tolerance, networking, and innovative thinking, respectively. These five factors are measured

through 130 paired statements. For this study, we created two distinct sets of paired statements from these original 130 pairs, with 30 pairs in each set. One set was presented in HMFCL format and the other in the traditional MFC format. Each set included statements representing each of the five factors in approximately equal measure, with statements representing each pole of each construct (i.e., positively and negatively worded statements). Statements in each pair were used exactly as they were in Futurenav Edge without any changes or modifications in the short version. Thus, pairings included statements that were matched in valence (positively worded versus negatively worded) and in social desirability ratings. In the HMFCL format, we added a 4-point Likert-type rating scale for each statement within a pair, as displayed in Table 3.

Another difference between the original PSQ and Futurenav Edge used in this study is that the original PSQ presents some statements in academic contexts, whereas Futurenav Edge references workplace contexts. For example, if a PSQ statement read, "I'm usually on time for class," we modified this statement to read, "I'm usually on time for work." Because these changes are contextual, they should not have a significant impact on construct measurement or conceptualization. In fact, the workplace contextualization represents an assessment version that pre-dates the PSQ (i.e., WorkFORCE Assessment for Job Fit; Naemi et al., 2014).

We also designed two surveys to assess participants' perceptions of the HMFCL and MFC formats. The first survey is an independent feedback survey in which participants responded to seven questions (five quantitative, two qualitative) designed to measure perceptions of the two response formats independently (see the appendices). The second survey is a comparative feedback survey with nine questions (eight quantitative, one qualitative) designed to capture participants' perceptions of the two response formats comparatively (see the appendices). In other words, the independent survey gathers participants' perceptions to the HMFCL and MFC formats independently of each other, whereas the comparative survey gathers perceptions of the two formats side by side. Quantitative questions utilized a Likert-type format, whereas qualitative questions utilized an open-ended or constructed-response format. We also included a brief background information questionnaire to tabulate participants' demographic data (e.g., age, race/ethnicity, gender) and employment

status/experience. Finally, we also captured participants' response time per item pairing for each of the HMFCL and MFC formats.

Sample

We recruited 53 participants (Table 6) through an online crowdsourcing platform (IntelliZoom).

Table 6. Participant Demographics

Demographic	*n*	%
Gender		
Male	22	41.5
Female	29	54.7
No response	2	3.8
Race/ethnicity		
American Indian or Alaskan Native	1	1.9
Asian	9	17.0
Black or African American	2	3.8
White	37	69.8
Some other race/ethnicity	2	3.8
No response	2	3.8
Age range (years)		
25–34	9	17.0
35–44	18	34.0
45–54	13	24.5
55+	11	20.8
No response	2	3.8
Current employment status		
Employed full-time	41	77.4
Employed part-time	9	17.0
Self-employed	1	1.9
No response	2	3.8
Current employment industry		
Education	11	20.8
Manufacturing	7	13.2
Information technology	5	9.4
Healthcare	4	7.5
Retail/wholesale trade	4	7.5
All other industries [a]	20	37.8
No response	2	3.8

Note. $N = 53$.
[a] "All other industries" is represented by 14 industries, each of which includes 1 to 3 participants.

The majority of participants self-described as female (54.7%) and as White (69.8%). All participants were over the age of 25 years, with 58.5% ranging in age from 35 to 54 years. Over three quarters (77.4%) of the sample was currently employed full-time. Nineteen workplace

industries were represented, with the most common including education (20.8%), manufacturing (13.2%), and information technology (9.4%).

Study Design

We utilized a within-subjects design for this study, as is commonplace in A/B user experience testing research. First, the two PSQ response format versions (HMFCL and MFC) were presented to each respondent in two blocks. Each block was followed by the independent feedback survey. These two blocks were counterbalanced such that roughly half of the respondents received the HMFCL block first, whereas the remaining participants received the MFC block first. Fifty-seven percent of respondents received the HMFCL block first. Following the two blocks, all participants received first the comparative feedback survey and then the background information questionnaire.

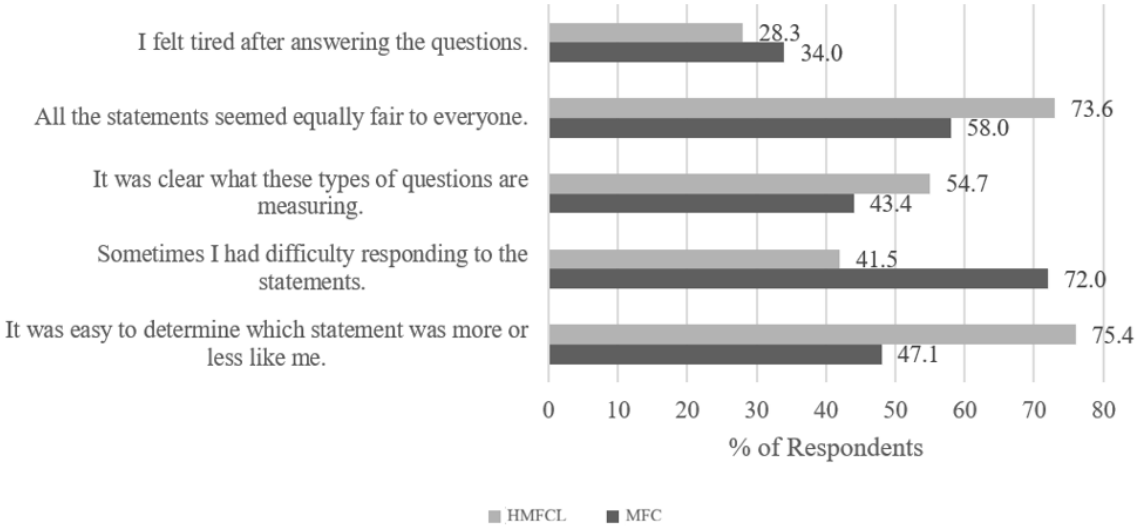
The survey included two short, open-ended questions. Thematic coding for the open-ended questions were conducted using a combination of human-coding and Microsoft Copilot (https://www.microsoft.com/microsoft-copilot). Copilot was used to brainstorm on potential high-level themes that were occurring within the responses. These themes were then used as the starting place for human coding, where they were then verified or modified by a human rater to identify the final common themes.

Results

Independent Feedback Survey

Due to the relatively small sample size, we emphasized descriptive and effect size interpretations over statistical significance. Figure 1 displays the results of the independent feedback surveys that immediately followed each of the HMFCL and MFC formats respectively.

Figure 1. Independent Feedback Survey Results



Note. HMFCL = hybrid multidimensional forced-choice Likert format; MFC = multidimensional forced-choice format. *N* = 53. Values represent the percentage of respondents who responded Agree or Strongly Agree on a 5-point scale.

To varying degrees, participants responded more favorably to the HMFCL format than the MFC format across all five of the quantitative items. The largest differences involved the two formats' ease of use. Specifically, participants were more likely to agree or strongly agree with the statement, "It was easy to determine which statement was more or less like me," for the HMFCL format (75.4%) than the MFC format (47.1%; $\chi^2 = 13.75$, $p < .01$, Cramer's $V^1 = 0.51$). Participants were also less likely to agree or strongly agree with the statement, "Sometimes I had difficulty responding to the statements," for the HMFCL format (41.5%) than the MFC format (71.7%; $\chi^2 = 8.64$, $p < .01$, $V = 0.28$). Differences were somewhat less pronounced for perceptions of fairness ("All the statements seemed equally fair to everyone"; 73.6% HMFCL versus 58.5% MFC; $\chi^2 = 2.06$, $p = .15$, $V = 0.14$), clarity/transparency ("It was clear what these types of questions are measuring"; 54.7% HMFCL versus 43.4% MFC; $\chi^2 = 0.94$, $p = .33$, $V = 0.09$), and fatigue ("I felt tired after answering the questions"; 28.3% HMFCL versus 34.0% MFC; $\chi^2 = 0.18$, $p = .67$, $V = 0.04$), but all favored the HMFCL format over MFC.

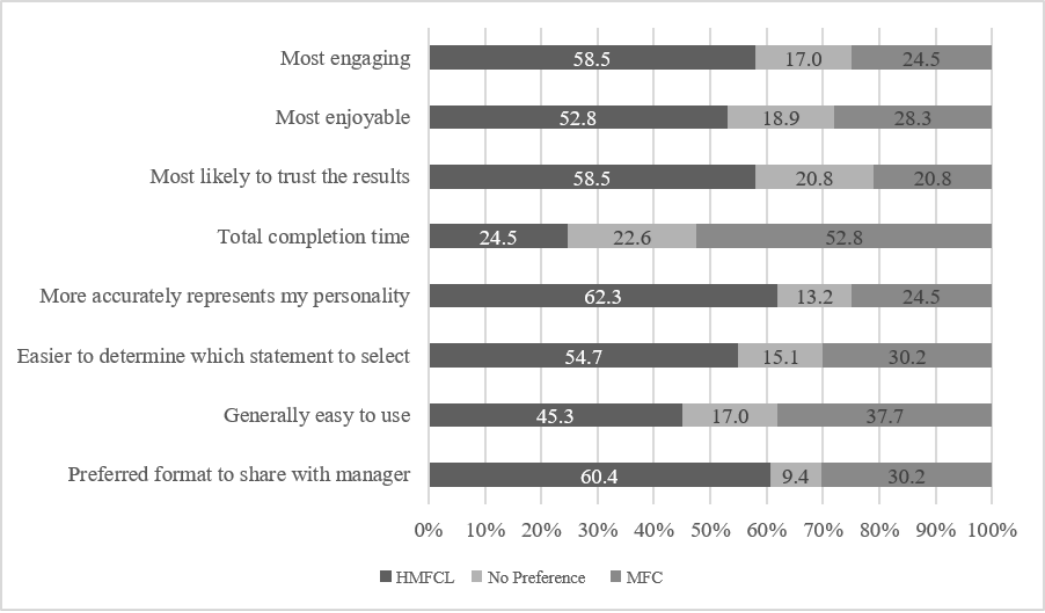
The independent feedback survey also included two open-ended questions for participants to describe what they liked most and least about the respective response formats. For the HMFCL format, participants most commonly liked the opportunity to self-reflect

(32.1%), having more response options (20.8%), and ease of responding (17.0%). For features that participants liked least about the HMFCL format, the most common responses included wanting to be able to select the same rating for both statements (22.6%), the overall length of the HMFCL version of the PSQ (15.1%), and that the statements were sometimes phrased too ambiguously (13.2%). For the MFC format, participants most commonly liked that the items were thought-provoking (32.1%) and that the format was easy and quick to use (30.2%). The most common feature that participants liked least about the MFC format was that they were forced to choose between the two options (54.7%). Sample responses are in Appendix B.

Comparative Feedback Survey

The comparative feedback survey included eight quantitative items in which participants were asked whether they preferred the HMFCL or MFC format across various metrics. As depicted in Figure 2, participants preferred the HMFCL format based on seven of eight metrics.

Figure 2. Comparative Feedback Survey Results



Note. HMFCL = hybrid forced-choice Likert format; MFC = multidimensional forced-choice format. *N* = 53. Values represent the percentage of respondents reporting a preference for the respective response formats.

The largest differences involved trustworthiness (58.5% HMFCL; 20.8% MFC; $\chi^2 = 14.24$, $p < .01$, V = 0.37), followed by the respective format being able to accurately represent individuals' personality (62.3% HMFCL; 24.5% MFC; $\chi^2 = 13.86$, $p < .01$, V = 0.36), engagement

level (58.5% HMFCL; 24.5% MFC; $\chi^2 = 11.22$, $p < .01$, V = 0.32), ease in selecting one statement over the other (54.7% HMFCL; 30.2% MFC; $\chi^2 = 5.56$, $p < .05$, V = 0.23), and enjoyability (52.8% HMFCL; 28.3% MFC; $\chi^2 = 5.63$, $p < .05$, V = 0.23).

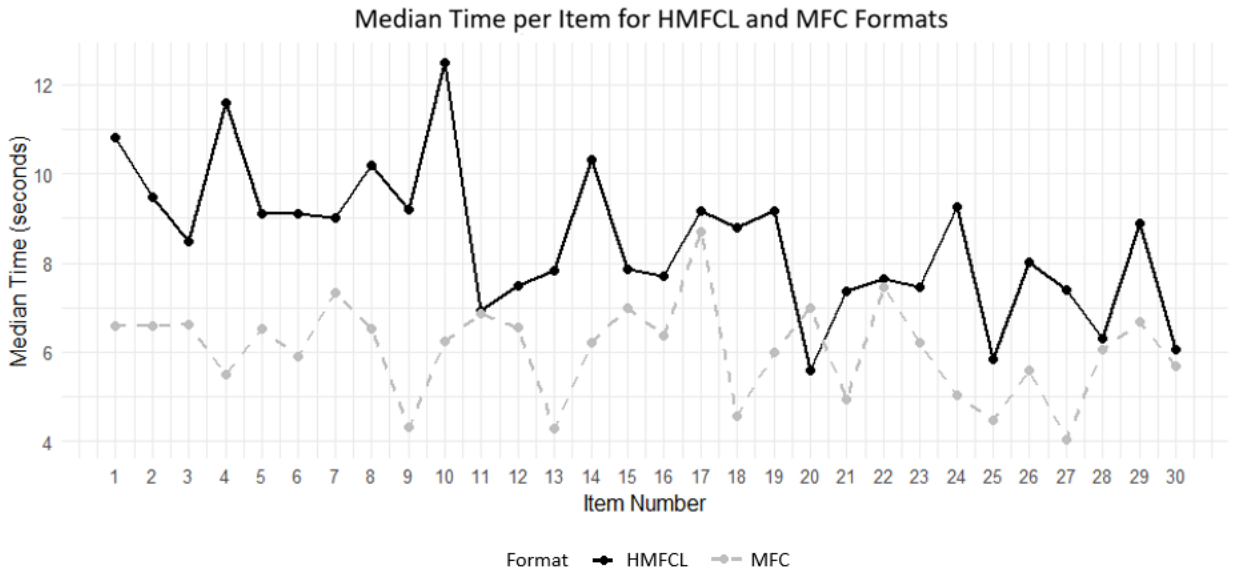
Overall, participants slightly preferred HMFCL in terms of its ease of use (45.3% HMFCL; 37.7% MFC; $\chi^2 = 0.35$, $p = 0.55$, V = 0.06). Participants also stated a strong preference for HMFCL if the assessment results were going to be shared with their manager (60.4% HMFCL; 30.2% MFC; $\chi^2 = 8.64$, $p < .01$, V = 0.28). The only metric that favored MFC was response speed for the overall assessment (24.5% HMFCL; 52.8% MFC; $\chi^2 = 7.80$, $p < .01$, V = 0.27).

The comparative feedback survey also included one qualitative (i.e., open-ended) question asking participants what they liked most about their preferred response format. Of the 32 participants who preferred the HMFCL format, the most common response (81.3%) involved liking the additional response choices relative to the MFC format. Of the 16 participants who preferred MFC, the most common response (68.8%) stated that they felt that MFC was easier and faster than HMFCL. Sample responses are displayed in Appendix B.

Timing Data

For each response format, we also analyzed response time data per item and for each set of 30 paired statements. At the item level, median response times were generally longer for the HMFCL format than the MFC format (Figure 3). However, this difference was more pronounced early in the assessment. By the time respondents reached the halfway mark of the 30 statements, these differences became almost negligible. This trend suggested that, with practice, respondents generally overcame any initial challenges with the unfamiliarity of the HMFCL format. Overall, the mean total response time was 5.53 minutes for the HMFCL format and 4.14 minutes for the MFC format (Table 7; $t = 7.18$, $p < .01$, Cohen's $d^2 = 1.31$). Together, the item- and assessment-level results supported participants' perceptions that the HMFCL format required a longer completion time.

Figure 3. Item-Level Response Time Data for Hybrid Multidimensional Forced-Choice Likert (HMFCL) and Multidimensional Forced-Choice (MFC) Formats



Note. N = 53. Each item represents one pair of statements.

Table 7. Assessment-Level Response Time Data (Minutes) for HMFCL and MFC Formats

Format	Median	Mean	*SD*	Min	Max
HMFCL	4.64	5.53	3.51	1.57	23.0
MFC	3.31	4.14	3.13	0.97	21.3

Note. HMFCL = hybrid multidimensional forced-choice Likert format; MFC = multidimensional forced-choice format. *N* = 53.

Discussion

The MFC response format represented a significant breakthrough in psychological assessment by addressing social desirability concerns inherent in the popular Likert response format. However, in our applications, negative feedback from test takers has hindered a more widespread adoption of MFC. Our HMFCL format was designed to address this feedback while maintaining MFC elements, and preliminary evidence suggests that it shows promise in this regard. By virtually every metric, participants preferred the HMFCL format over MFC. Participants generally found HMFCL to be easier to respond to and more accurate, fair, engaging, and trustworthy than MFC. The primary drawback to the HMFCL was a slightly longer completion time relative to MFC, though this issue appeared to become less pronounced as participants became accustomed to the HMFCL format. Providing clearer examples and practice

items prior to beginning the assessment may further facilitate shorter response times. Some participants still disliked the forced-choice nature of HMFCL, but removing it would defeat a significant and necessary feature of this format. Overall, although there were still some criticisms of the HMFCL method, they paled in comparison to participants' dislike of the MFC format's restrictiveness, which was the most common open-ended feedback we received across both formats. These MFC criticisms echo those found in previous studies (e.g., Borman et al., 2024; Dalal et al., 2021). In fact, much of the negative feedback mentioned in the context of the HMFCL format was actually referencing elements that were independent of HMFCL (e.g., statement content).

The novelty of the HMFCL format suggests myriad opportunities for future research. Perhaps most important is to identify an ideal scoring model. The HMFCL format is amenable to traditional MFC scoring models such as the generalized graded unfolding model (Roberts et al., 2000) and the multi-unidimensional pairwise preference two-parameter logistic model (MUPP-2PL; Fu, Tan, & Kyllonen, 2024) by simply recoding the items dichotomously prior to scoring (e.g., *very like me* = 1; *like me* = 0). However, scoring models used for response formats such as MFC and graded response are suboptimal because they only capture relative response preferences. HMFCL includes relative and absolute response data, requiring additional parameters in the scoring model. Moreover, by dichotomizing HMFCL response data, MFC scoring models remove valuable response variations collected through HMFCL. Conversely, a new scoring model would leverage the additional data provided by HMFCL. For example, a new model is necessary to differentiate between a respondent who selects Very Like Me and Like Me from another respondent who selects Unlike Me and Very Unlike Me on the same pair of statements. Traditional MFC and graded-response scoring models would consider these two participants to be identical because their respective response patterns demonstrate the same relative preference. However, these participants display very different absolute levels of the constructs being measured and, in fact, are of opposite valence. We are currently developing a model to capture these distinctions along with all other possible response patterns that may be captured by HMFCL. In the interim, we have applied MFC scoring models (e.g., MUPP-2PL) to HMFCL data.

Once the scoring model is identified, additional fundamental psychometric evaluations may be conducted. This new scoring model is necessary to accurately calculate characteristics such as reliability (e.g., internal consistency) and fairness (e.g., differential item functioning). Adequate scoring will also facilitate various validity analyses such as construct validity correlations with alternative measures. This scoring model will also assist in investigating the fake-resistance of HMFCL. This characteristic is of critical importance as it represents the primary motivation for creating the MFC format, whose elements are maintained within HMFCL. These studies would benefit from longer assessments and larger, more demographically diverse participant samples. For example, item-level analyses could determine if HMFCL generates additional cognitive load for reverse-scored items. Although we deliberately omitted a neutral option in our HMFCL demonstration, subsequent applications may evaluate its potential usefulness. HMFCL may also be expanded to additional statement groupings by extending the number of possible responses. For example, sets of three statements may be measured using six Likert options (*very like me, like me, slightly like me,* etc.) instead of four. These additional options would also permit the measurement of variation in relative preferences across respondents for item pairs, similar to the graded-preference format. These studies could provide a more in-depth investigation of the benefits and potential limitations of the HMFCL's ability to capture absolute construct values, such as social desirability influences as well as centrality or extremity bias.

Future research could address limitations in our current study. For example, these studies should also utilize larger, more diverse nonconvenience samples than our study to increase generalizability and precision of analytical results. Future research could also follow up on our survey results in more detail through methods such as interviews or cognitive lab studies (e.g., how respondents interpreted the fairness of the response formats). Direct comparisons between HMFCL and other formats such as graded pairs could also be informative. Collectively, these studies should facilitate the use of HMFCL in research settings and in higher-stakes applications.

References

Bech, M., Gyrd-Hansen, D., Kjær, T., Lauridsen, J., & Sørensen, J. (2007). Graded pairs comparison - does strength of preference matter? Analysis of preferences for specialised nurse home visits for pain management. *Health Economics*, *16*(5), 513–529. https://doi.org/10.1002/hec.1159

Borman, T. C., Dunlop, P. D., Gagné, M., & Neale, M. (2024). Improving reactions to forced-choice personality measures in simulated job application contexts through the satisfaction of psychological needs. *Journal of Business and Psychology*, *39*(1), 1–18. https://doi.org/10.1007/s10869-023-09876-w

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*(3), 460–502. https://doi.org/10.1177/0013164410375112

Brown, A., & Maydeu-Olivares, A. (2018). Ordinal factor analysis of graded-preference questionnaire data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 516–529. https://doi.org/10.1080/10705511.2017.1392247

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes contexts. *Journal of Applied Psychology*, *104*(11), 1347–1368. https://doi.org/10.1037/apl0000414

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*(3), 267–307. https://doi.org/10.1207/s15327043hup1803_4

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Dalal, D. K., Zhu, X. S., Rangel, B., Boyce, A. S., & Lobene, E. (2021). Improving applicant reactions to forced-choice personality measurement: Interventions to reduce threats to test takers' self-concepts. *Journal of Business and Psychology*, *36*, 55–70. https://doi.org/10.1007/s10869-019-09655-6

Dueber, D. M., Toland, M. D., Lingat, J. E., Love, A. M. A., Qiu, C., Wu, R., & Brown, A.V. (2022). To reverse item orientation or not to reverse item orientation, that is the question. *Assessment, 29*(7), 1422–1440. https://doi.org/10.1177/10731911211017635

Fu, J., Kyllonen, P. C., & Tan, X. (2024). From Likert to forced choice: Statement parameter invariance and context effects in personality assessment. *Measurement: Interdisciplinary Research and Perspectives, 22*(3), 280–296. https://doi.org/10.1080/15366367.2023.2258482

Fu, J., Tan, X., & Kyllonen, P. C. (2024). Information functions of Rank-2PL models for forced-choice questionnaires. *Journal of Educational Measurement, 61*(1), 125–149. https://doi.org/10.1111/jedm.12379

Hayes, T. L. (2007). Review of *A closer examination of applicant faking behavior* [Review of the book *A closer examination of applicant faking behavior*, by R. L. Griffith & M. H. Peterson, Eds.]. *Personnel Psychology*, *60*(2), 511–514. https://doi.org/10.1111/j.1744-6570.2007.00081_5.x

İlhan, M., Güler, N., Taşdelen Teker, G., & Ergenekon, Ö (2024). The effects of reverse items on psychometric properties and respondents' scale scores according to different item reversal strategies. *International Journal of Assessment Tools in Education, 11*(1), 20–38. https://doi.org/10.21449/ijate.1345549

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, *13*(4), 371–388. https://doi.org/10.1207/S15327043HUP1304_3

Krammer, G. (2020). Applicant faking of personality inventories in college admission: Applicants' shift from honest responses is unsystematic and related to the perceived relevance for the profession. *Journal of Personality Assessment, 102*(6), 758–769. https://doi.org/10.1080/00223891.2019.1644342

Lingel, H., Bürkner, P.-C., Melchers, K. G., & Schulte, N. (2024). Measuring personality when stakes are high: Are graded paired comparisons a more reliable alternative to traditional forced-choice methods? *Organizational Research Methods*, *28*(4), 610–640. https://doi.org/10.1177/10944281241279790

Naemi, B., Seybert, J., Robbins, S., & Kyllonen, P. (2014). *Examining the WorkFORCE™ Assessment for Job Fit and core capabilities of FACETS™* (ETS Research Report No. RR–14-32). ETS. https://doi.org/10.1002/ets2.12040

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3–32. https://doi.org/10.1177/01466216000241001

Schulte, N., Kaup, L., Bürkner, P.-C., & Holling, H. (2024). The fakeability of personality measurement with graded paired comparisons. *Journal of Business Psychology, 39*(5)*,* 1067–1084. https://doi.org/10.1007/s10869-024-09931-0

Williams, K. M. (2022). The Occupational Performance Assessment–Response Distortion (OPerA-RD) scale: Measuring bias in performance evaluation. *Journal of Personnel Psychology, 21*(4), 185–196. https://doi.org/10.1027/1866-5888/a000301

Zhang, B., Luo, J., & Li, J. (2024). Moving beyond Likert and traditional forced-choice scales: A comprehensive investigation of the graded forced-choice format. *Multivariate Behavioral Research, 59*(3), 434–460. https://doi.org/10.1080/00273171.2023.2235682

Appendix A. Feedback Surveys

The following feedback surveys were developed by the study authors for use in this study.

Independent Feedback Survey

Note. Participants received this survey twice, once after responding to each assessment section (hybrid multidimensional forced-choice Likert [HMFCL] and multidimensional forced-choice [MFC]).

Instructions: On a scale from 1 (*strongly disagree*) to 5 (*strongly agree*), please use the following items to rate your reaction to the previous questions:

1. I felt tired after answering the questions.
2. All of the statements seemed equally fair to everyone.
3. It was clear what these types of questions were measuring.
4. Sometimes I had difficulty responding to the statements.
5. It was easy to determine which statement was more or less like me.

For the following two questions, please type your response:

1. What did you like most about the previous questions?
2. What did you like least about the previous questions?

Comparative Feedback Survey

Instructions: Based on the following factors, please select whether you preferred format A (HMFCL), format B (MFC), or had no preference:

(a) Most engaging

(b) Most enjoyable

(c) Most likely to trust the results

(d) Total amount of time to complete all questions

(e) More accurately represents my personality

(f) Easier to determine which statement to select

(g) Generally, easy to use

(h) If you had to complete questions like this as part of your current job and your manager was going to see the results, which format would you prefer?

For the following question, please type your answer:

(i) What did you like most about the format that you preferred?

Appendix B. Feedback Survey Results

Table B1. Sample Open-Ended Responses (Independent Feedback Survey)

HMFCL	
Liked most	Liked least
Chance to self-reflect (32.1%): • "The statements feel relevant to me and they are thought provoking." • "It got me to think about my behaviors, personality, and attitudes."	Wanted to select the same rating for both responses (22.6%): • "Some of them had situations where I wanted to select the same rating for both, but I couldn't by design" • "I didn't like that I couldn't give the same answer to each question in a pair."
Having more options (20.8%): • "It gave more options on which statement was most like you" • "There had better middle ground answers and some nuance compared to the previous list of questions"	Too long (15.1%): • "Just way too long to maintain close focus" • "30 questions seems like a bit much"
Easy to understand and answer (17.0%): • "Additionally, the questions were clear and well-formulated, making it easy for me to understand what was being asked and to provide an accurate response." • "It was easier to answer this way"	Questions were ambiguous/complex (13.2%): • "The questions sometimes feel one sided. Neither would apply to me but I need to figure them out." • "Some questions were still written weirdly, so it was not always easy to understand how to answer"
MFC	
Liked most	Liked least
Thought-provoking (32.1%): • "Got me think about issues at work and in personal life" • "They were interesting questions, made you think."	Didn't like choosing between the two options (54.7%): • "Again, it was hard to compare certain things. Some things were both very unlike me, but I HAD TO pick one. Sometimes it almost felt like lying because I had to choose answers that were not like me, but the answer was just a smidge closer than the other one." • "Some of the statements are against each other at the extreme that I don't think I can pick one fairly."
Easy/quick to use (30.2%): • "I like that the statements weren't too long and they were quick and easy to read" • "They were easy to understand even if it didn't apply to me"	

Note. HMFCL = hybrid multidimensional forced-choice Likert format; MFC = multidimensional forced-choice format. *N* = 53.

Table B2. Sample Open-Ended Responses (Comparative Feedback Survey)

Preferred HMFCL (*N* = 32)	Preferred MFC (*N* = 16)
Liked having additional choices to increase accuracy (81.3%): • "the answers are more forgiving with Format A [hybrid], Format B [paired] was oftentimes too absolute" • "I didn't feel the Format B [paired] questions were an accurate representation of my personality	Felt MFC was easier and faster (68.8%): • "It's a bit more straightforward, each question requires less cognitive focus" [preference for paired] • "I like that the descriptions were simplified and only two options to choose, rather than 4"

Note. HMFCL = hybrid multidimensional forced-choice Likert format; MFC = multidimensional forced-choice format. Values represent percentage of respondents within preference group.

Notes

[1] Guidelines for interpreting Cramer's V effect size are small ≈ ±0.10, medium ≈ ±0.30, and large ≈ ±0.50 (Cohen, 1988).

[2] Guidelines for interpreting Cohen's *d* effect size are small ≈ ±0.10, medium ≈ ±0.30, and large ≈ ±0.50 (Cohen, 1988).

www.ets.org

Copyright © 2026 by ETS. The Eight-Point logo is a trademark of ETS. All other trademarks are property of their respective owners.