

# Recruiting a Person-Centered Direct Care Workforce Through Situational Judgement Tests: A Pilot Study in the Community Support of Older People in England

Mark Wilberforce, BSc, MSc, PhD,<sup>1,\*</sup> Abby Dunn, MA, MRes, PhD,<sup>2</sup> and Paul A. Tiffin, MD, MBBS, BMedSci(Hons) FRCpsych FHEA<sup>3,4</sup>

<sup>1</sup>School of Business & Society, University of York, York, UK.

<sup>2</sup>Family Fund, York, UK.

<sup>3</sup>Department of Health Sciences, University of York, York, UK.

<sup>4</sup>Hull York Medical School, York, UK.

\*Address correspondence to: Mark Wilberforce, BSc, MSc, PhD. E-mail: [mark.wilberforce@york.ac.uk](mailto:mark.wilberforce@york.ac.uk)

Decision Editor: Howard B. Degenholtz, PhD, FGSA

## Abstract

**Background and Objectives:** Studies have found that older people value care workers' character traits and interpersonal judgment even more highly than their technical skills. Yet identifying these traits at recruitment is challenging. This study aimed to evaluate the first situational judgement tests (SJTs) for direct care workers.

**Research Design and Methods:** Online tests were conducted with 251 care workers and members of the public in England. Participants evaluated the appropriateness of 61 potential behavioral responses to 11 "critical incidents," each depicting challenging care work scenarios. Data collection included a measure of personality traits. A subsample of 72 participants completed a second "test-retest" assessment.

**Results:** A majority of test-takers (53%) found the test easy/very easy to complete, and most (55%) participants who worked in care reported the scenarios were highly realistic. Psychometric tests were positive. Test scores were unidimensional under a Confirmatory Factor Analysis (RMSEA = 0.038), and items broadly fit Rasch assumptions. Test-retest reliability ( $\rho = 0.77$ ) was acceptable, and for the general public sample, a modest increase in perceptions of the social standing of care work was observed. Test scores were positively correlated with 2 personality traits: agreeableness ( $r = 0.250$ ,  $p < 0.001$ ) and openness to experience ( $r = 0.179$ ,  $p = .005$ ). Test scores were not related to age, gender, or education level.

**Discussion and Implications:** The findings indicate support for the use of SJTs in direct care work. Its psychometric properties appear satisfactory, and collectively give confidence in the use of SJTs for assessing the suitability of candidates during recruitment. Further research should corroborate these findings in a new sample, and examine the relationship between test scores and job performance.

**Translational Significance:** The use of situational judgement tests in direct care work adds a crucial objective component to the "art" of recruitment. Even structured interviews are known to be subject to bias, and objective assessments can support decision-making. The tests can identify individuals with the requisite person-centered values that are highly desired by older people receiving care at home. The study finds the tests to be psychometrically robust, acceptable to test-takers, and correlated with key personality traits essential for care work. A new freely available platform, [www.curiousaboutcare.org.uk](http://www.curiousaboutcare.org.uk), implements these tests for recruitment.

**Keywords:** Long-term care, Recruitment, Situational judgement tests, Values-based recruitment, Workforce

The growing number of older people with care needs is forecast to bring new demands on community long-term care (defined broadly as paid assistance in everyday home living for reasons of disability or frailty). In most nations, the additional pressures on care systems are already outpacing growth in the workforce supply of personal care aides and support staff (hereafter, "care workers"); a trend that is likely to continue in coming decades (OECD, 2020; Spetz et al., 2019). Although different countries face distinct challenges

and contexts, a familiar narrative is common to most. Care work lacks a coherent career structure with pay progression based on experience; has unclear expectations about requisite knowledge; has a poor social standing relative to health-care employees with equivalent skills, and has employment conditions characterized by job instability and casualization, including temporary or zero-hours contracts (OECD, 2020). Turnover rates are high, with average tenure in long-term care being two years lower than for the labor force as a whole

Received: March 29 2023; Editorial Decision Date: September 10 2023.

© The Author(s) 2023. Published by Oxford University Press on behalf of The Gerontological Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

(Llena-Nozal et al., 2022), with likely detrimental consequences for care outcomes (Antwi & Bowblis, 2018). To meet chronic workforce shortages and future demand, national governments and care sector bodies seek to bring in new workers from other segments of the labor force. Examples include policies to attract ex-offenders, military veterans, the long-term unemployed, and other under-represented groups into care work (Manthorpe & Bramley, 2020; Ministry of Justice, 2021; OECD, 2020). In most developed nations the sector relies heavily on immigrant populations to meet labor needs, with growing attention paid to policies that ease labor flows (Zallman et al., 2019).

Care work is also characterized by a “skills mismatch” (Llena-Nozal et al., 2022). Although there are few formal skill requirements, minimum education, or skills licenses needed to take up a care worker post in most countries, the lived reality of care work is one which goes far beyond simple tasks. Support for older people at home with complex needs, particularly linked to dementia or end-of-life care, demands a substantial range of judgments and interpersonal skills that are negotiated with (sometimes reluctant or even hostile) clients, under stark time pressures (Abrams et al., 2019; Leverton et al., 2021). Moreover, care work equates to more than the functional activities that support daily living. Workers commonly provide emotional and social support, through companionship and relational work, while managing complex boundaries in the context of heavy emotional labor (D'Astous et al., 2019; Scales, 2021). Care workers frequently manage confrontational behaviors and ethical dilemmas placing them under considerable pressure (Delp et al., 2010) and yet often work with little supervision or formal structured working environments. Pay and terms of employment are not commensurate with these demands, with recruiting employers relying more on “intrinsic” rewards from meaningful, intimate, and reciprocal relationships available in care work that cannot be easily attained from other service occupations (Hussein, 2017; Morgan et al., 2013).

Identifying candidates well-suited to care work can be challenging. In reviewing evidence on what service users and family carers seek in care, personal character, and values appear to be more important than technical skills and knowledge. In general, “values” can be understood to be enduring beliefs that steer people toward exhibiting certain behavioral dispositions. Empirically, they are correlated with personality traits (Roccas et al., 2002). However, in contrast to personality traits, values tend to be viewed as more conscious, with associated cognitive frameworks for guiding actions in response to different stimuli or situational dilemmas. Previously, Motowidlo et al. (2006) have also referred to “implicit trait policies” (ITPs), which are implicit beliefs about causal relations between personality traits and behavioral effectiveness. For example, if an employee believes that agreeableness is important in relation to carrying out one's duties they are more likely to exhibit agreeable behavior when a situation arises that gives opportunity to demonstrate this trait.

Values are relatively well understood in health and allied professions because they are learned and instilled during training and normalized by ongoing socio-occupational processes. By contrast, care worker values are less transparent or clearly defined. Qualitative studies imply that desired values are those that have developed naturally as part of everyday human relations, rather than those instilled by formal structures. Manthorpe et al., (2017)'s interviews with 60 care

recipients and families even found a widespread view that formal training could *undermine* values, because participants expressed the importance of “genuine” traits over those imposed externally. Moreover, given the degree of independent working and limited supervision, the paucity of training opportunities, and the lack of socialization through which care work standards can propagate, chances to learn and acquire values are limited.

Consequently, promoting values in care work is likely to depend upon effective recruitment processes more heavily than for other occupations. Interviews are the most common form of approach to selection, but on their own are subject to substantial biases, particularly when assessing “softer” skills (CIPD, 2015). Even highly structured interviews are influenced by perceptions of candidates formed from outside the job requirements (Barrick et al., 2012), and decision-making can be improved by objective assessments alongside interviews (Hoffman et al., 2018).

Among objective methods with some promise are “situational judgement tests” (SJTs). SJTs involve candidates reviewing (usually online) written, animated, or filmed job-based scenarios in which chosen values are depicted as being put to the test. After reviewing the scenario, candidates appraise a number of possible behavioral responses. This process is repeated across multiple questions linked to several scenarios. The candidate's responses are rated according to a previously developed scoring key. In this way, employers can evaluate the extent to which candidates understand key values, as depicted in the scenarios and response options. Thus, SJTs, in this situation, can be considered a special case of a “procedural knowledge test” (Tiffin et al., 2020).

The use of SJTs as a component of personnel selection is widespread in many occupations and industries, including the public sector (e.g., policing, teaching, and medicine). Their principal benefit is in providing a cost-effective source of relatively objective, quantitative data which can be used alongside or independently of other selection assessments, such as interview-based processes. The validity of the scores produced by SJTs tends to be evaluated by assessing their ability to predict construct-relevant measures of on-the-job interpersonal functioning. In this regard there is good meta-analytical evidence for the overall validity of SJT scores, both in personnel selection in general (McDaniel et al., 2007), and specifically in medicine (Webster et al., 2020). Indeed, in some contexts face-to-face processes add little or no incremental predictive validity, over and above that provided by a well-designed SJT (Davison et al., 2016).

However, the use of SJTs is not common in long-term care, with only proprietary tests available at high costs, which are often prohibitive for a sector hampered by financial austerity. A previous paper (Dunn et al., 2023) describes the development of specific long-term care SJTs that are available freely to care worker employers. The SJTs were developed via a multi-stage research study, and designed to test situated judgment in community care for older people. The three concepts selected for a content generation were identified as central to person-centered care (Wilberforce et al., 2017): *understanding the person*; *supporting involvement in decisions*; and *reciprocal care relationships*. Scenarios were based on the “critical incident” interviewing approach with community care workers in England. Interviews sought examples of each value being exhibited through real-life care experiences. A total of 11 scenarios were used in the

test, with candidates asked to rate the appropriateness of 61 behavioral responses on a short Likert scale. An overview of the 11 scenarios is presented in Table 1. It is usual practice for SJT scoring rubrics to be set by “subject matter experts,” most commonly senior professionals working in the field. However, uniquely in terms of SJT development, the scoring key was formed by older people with lived experience, using a modified Delphi (Dunn et al., 2023). This was considered of key importance given the “person-centered care” focus of the tool.

The psychometric properties of this SJT are, as yet, untested. It is not known whether the test generates scores that can be considered reliable and valid. Further, differences in test scores between different categories of test-takers have not been examined, such as whether there are differences by education level. Finally, no evidence is currently available on the acceptability of the SJTs to potential test-takers. More formally, this paper aims to (a) examine the psychometric properties of the SJTs, to include an assessment of structural dimensionality, acceptability, fidelity, internal and test/retest reliability, and convergent validity; and (b) to undertake an exploratory analysis to understand variation in SJT scores and the likely impact of administration on perceptions of care work.

## Method

### Instrument and Measures

The SJT was implemented within Qualtrics as a self-completed online survey. Each scenario was introduced at the head of the page, and each behavioral response was found beneath, which cycled through to the next scenario. For each of 61 behavioral responses, participants were asked to rate “how good or bad would the following actions be?” on a 4-point “very good/ good/ bad/ very bad” Likert-type scale. In applying the subject-matter expert scoring system (see Dunn et al., 2023), responses were effectively collapsed into a

binary system: participants scored one point if their response was on the same side of the midpoint as required by the scoring rubric (and zero otherwise). The SJT was accompanied by questions on socio-demographic characteristics, and the Big Five Inventory 30-item version (Soto & John, 2017), to measure self-reported personality traits. As a measure of the acceptability of the SJTs, a question was also included on the perceived difficulty of the test.

Two samples were recruited (described below): those currently employed as care workers (“incumbents”); and a general population sample without care experience (“care naïve”). Questionnaire routing allowed each to be asked additional questions. As a measure of “fidelity,” the former were asked about how closely the SJTs “reflect the reality of care work” based on their own experiences with the job (3 point Likert scale: very much.../ somewhat.../ does not ... ).

As a measure of care perceptions, the latter, care naïve, sample were asked about their attitudes toward care work using five statements designed by the research team: “Care work is a low-skill job,” “Care workers should be proud of their job,” “Care work is emotionally rewarding,” “Care work involves making lots of decisions,” and “Care work is the same every day.” Participants’ agreement/disagreement was recorded using a 5-point “Strongly disagree – Strongly agree” Likert scale. These were asked before and after the completion of the SJTs.

### Sample

“Incumbent” care workers were recruited through home care providers across England. Advertisements were circulated through the UK Homecare Association (a national trade body) asking care providers interested to get in touch with the research team. Among interested providers, an information sheet was forwarded to care workers within the organization inviting them to participate in the online survey. No restrictions on hours worked or length of time in care were applied.

**Table 1.** An Overview of 11 Scenarios Forming the Situational Judgement Test.

Name	Critical incident
Lázsló	Lazslo, an outwardly friendly gentleman who has memory troubles, is in a café, and starts talking to people who he thinks he knows, but they do not know him.
Mrs. Haddow	Mrs. Haddow loves to knit, but her arthritis is meaning she is struggling to finish her scarf. This is getting her down.
Lloyd	Lloyd is a “people person,” and is supported primarily due to occasional loneliness. He notices that the care worker is not their usual self, and asks if there is anything that matter.
Stephen	Stephen has already got himself dressed and wants to sit and have a natter with care worker. However, family member objects, saying care worker is not paid to drink tea.
Ken	Ken is a gentleman who “lives life to the full.” He has diabetes. He wants care worker to order an (unhealthy) meal from a takeaway.
Mr. Singh	Care worker overhears Mr. Singh talking with a friend, and referring to the care worker as “being like a servant.”
Derek	Derek misplaces an item, and accuses care worker of theft. Care worker feels hurt.
Miss Holtby	Miss Holtby enjoys baking, but has been unable to do so after a fall, and lost confidence. She now feels she has recovered. At the end of the care worker’s visit, Miss Holtby intends to start baking.
Vera	Vera has quite advanced dementia. She is someone who has always taken pride in her smart appearance. She is in need of a bath, and her personal hygiene is suffering. However, she refuses to have a bath because she is enjoying being with “her baby” (a doll)
Mrs. Gupta	Mrs. Gupta is very angry that the care worker has brought a beef sandwich into her home for lunch, because of her religious views.
Margaret	Margaret lives alone after being widowed, and has quite advanced dementia. The care worker arrives and finds Margaret searching for her deceased husband.

“Care naïve” participants were recruited from the “Prolific” research database ([www.prolific.co](http://www.prolific.co)) as a general population research sample. The survey was open to all potential participants living in England, except that stratification took place on economic status to match the characteristics of new entrants to care work. During the study, a lack of representation from black and Asian participants was noted, which was rectified through a “booster” sample, through a time-limited social media advertising targeting minority groups. All “care naïve” respondents were also invited to take part in a second “retest” questionnaire two weeks after completion of the first. Ethical approval was granted from the University of York’s Social Policy and Social Work Ethics Committee (SPSW/S/20/6).

## Analysis

The analytical approach inspected multiple aspects of SJT functioning and properties, summarized in Table 2. The psychometric procedures broadly follow those outlined in Tiffin et al (2020). Dimensionality assesses whether the SJT measures a single construct as intended. The degree to which the responses from the SJT showed a unidimensional pattern was evaluated via Confirmatory Factor Analysis (CFA) adapted for the binary nature of the scores, using WLSMV (weighted least square mean and variance-adjusted) estimation within the Mplus 8.7 software package. In this way the fit statistics, including RMSEA (ideally < 0.05), CFI, and TFI (ideally > 0.90) could be examined after estimating a model whereby all the item responses were explained by a single unitary factor. A secondary test of unidimensionality was to conduct a principal components analysis of residuals after Rasch analysis.

Rasch analysis was conducted in Winsteps. Where score patterns conform to Rasch assumptions then the scores can be used as interval-level measurements (Rasch, 1960). Rasch analysis yields unique fit indices that are helpful for diagnosing measurement issues. These are “infit” and “outfit” values. These reflect the degree of conformity to the Rasch model (i.e. “Guttman sequence” observed). In this respect values

between 0.5 and 1.5 are considered ideal for measurement, <0.5 suggests the presence of redundant items whilst values > 2.0 are considered “degrading” for measurement (Linacre, 2003). Internal consistency was assessed through two test statistics. Cronbach’s alpha is a “classical” psychometric metric of internal reliability consistency and effectively reports an average split-half reliability value (i.e. the average inter-correlation among the item responses). The Person Separation Index is a Rasch, item response theory-based metric of test reliability (“test information”) and evaluates the extent to which distinct strata of test-takers can be reliably discriminated. That is, more “reliable” tests yield more information on each test-taker which can be used to discriminate between individuals.

Evidence for construct validity, in terms of convergent validity, was assessed through the evaluation of the degree of correlation between the SJT and the “Big Five” personality trait scores. The a priori hypothesis was that “agreeableness” would be positively and moderately correlated with the SJT score, because this incorporates compassion, respectfulness, and trust in the BFI-2-S, which are all traits that are well-aligned with person-centered care. Internal consistency was evaluated through Cronbach alpha (values > 0.7 considered conducive to measurement) and the Person Separation Index (values > 2.0 considered ideal). Test-retest reliability is evaluated by a Spearman correlation between test scores at two-time points, with a correlation > 0.7 considered “acceptable” (Nunnally, 1978).

## Findings

Sample characteristics for 251 respondents are presented in Table 3. The incumbent care worker sample appeared relatively representative of national data on care workers. For example, the research sample was 84% female (compared with 83% nationally) and 29% aged over 50 (compared with 27% aged 55 or over nationally; Skills for Care, 2022). The care naïve sample had a noticeably different age and ethnicity profile, being younger on average, and with fewer Black or Asian participants within their number.

**Table 2.** Overview of Analytical Framework

Framework	Description
Acceptability	Measured by self-reported perception of ease/difficulty in completing the SJTs
Fidelity	Measured through perception of the extent to which the SJTs represent real-life practice (incumbent care worker sample only)
Dimensionality	The extent to which the items can be assumed to measure a single construct. This was examined by assessing the degree of fit of the item responses to a single factor model on (binary) confirmatory factor analysis (CFA), supplemented with a Principal Component Analysis (PCA) of residuals after a Rasch analysis
Internal consistency	Measured by Cronbach alpha and the Person Separation Index derived from the Rasch analysis
Interval-level measurement	Whether SJT scores can be treated as an interval-level measure is evaluated by Rasch analysis in Winsteps, and item-level assessment of “infit” and “outfit”
Construct (concurrent) validity	Assessed through the correlation coefficients between the SJT and BFI-2-S scores
Test-retest reliability	Measured by the Spearman correlation between the first SJT score, and a second repeated administration two weeks later
Care perceptions	Measured through a bespoke scale formed of five items before and after completing the SJTs, with changes evaluated through a <i>t</i> -test (care naïve sample only)
Exploratory analysis	Examining variation in SJT score by socio-demographic variables using mix of <i>t</i> -tests and ANOVA with post hoc corrections

Notes: ANOVA = Analysis of Variance; BFI-2-S = Big Five Inventory-2-Short Form; SJT = Situational Judgement Test.



**Table 3.** Sample Characteristics.

Characteristics	Incumbent care workers ( <i>n</i> = 160) <i>n</i> (%)	Care naïve sample ( <i>n</i> = 90) <i>n</i> (%)
Gender		
Female	134 (84%)	70 (78%)
Male	26 (16%)	18 (20%)
Other	0 (0%)	2 (2%)
Age		
<30	29 (18%)	59 (66%)
31–50	85 (53%)	23 (26%)
51+	46 (29%)	8 (9%)
Highest qualification		
Higher education	85 (53%)	50 (56%)
A level or equivalent	21 (13%)	25 (28%)
GCSE or equivalent	32 (20%)	10 (11%)
None/ below GCSE	16 (14%)	4 (4%)
Ethnicity		
Asian/Asian British	22 (14%)	2 (2%)
Black/Black British	99 (23%)	1 (1%)
White	94 (62%)	59 (90%)
Mixed and other	3 (2%)	6 (7%)
Disability		
Yes, has a disability	8 (5%)	12 (13%)
Not disabled	148 (93)	78 (87%)
Not stated	4 (3%)	0 (0%)

Note: GCSE = General Certificate of Secondary Education.

### Acceptability and Fidelity

All participants were asked how easy or difficult they found the SJT to complete. Most participants (53%) found the questionnaire “easy” or “very easy” to complete, and a further 31% (*n* = 77) found it “neither easy nor difficult” to complete. Just 16% found the SJT “somewhat difficult” to complete and no respondent found it “very difficult” to complete. Care workers found the SJTs easier to complete than non-care workers.

As a test of fidelity, incumbent care workers were asked if the SJT reflected their own experience of care work: 55% (*n* = 78) felt it “very much” reflected care work and 44% (*n* = 62) felt it “somewhat” reflected care work. Just one person felt it did not reflect care work.

### Dimensionality

As outlined earlier, responses to each SJT item were scored according to a rubric set by people with lived experience (see Dunn et al., 2023). A one-factor CFA was conducted using robust weighted least squares as the estimator. The fit indices indicated that it was reasonable to assume that the response structure was unidimensional, adjudged by a CFI value of 0.917 and TFI value of 0.914 (values > 0.90 are taken as acceptable fit). The RMSEA was estimated to be 0.038 [90% confidence intervals: 0.034, 0.042]. The PCA of residuals after Rasch analysis supported a broadly unidimensional solution. Beyond the primary Rasch dimension, the second, third, and fourth contrasts accounted for a relatively small proportion of common variance suggesting any additional dimensions

would be relatively minor. The full results are published as [Supplementary Material](#).

### Internal Consistency

The items were collectively of adequate internal consistency, with a Cronbach’s alpha value of 0.85 from 61 items. The person separation index was 1.52.

### Interval-Level Measurement

A further feature of a potentially effective psychological measurement tool is that scores can be transformed to form an “interval” metric with a common unit of measurement. For this analysis, three items had to be excluded because over 99% of respondents had answered these items correctly so there was virtually no variance (and hence no information yielded). Of the 59 remaining response items, almost all showed good or acceptable fit to Guttman patterns expected within a Rasch analysis (see [Supplementary File](#)). No items showed an infit value exceeding 2.0. Outfit values exceeded 2.0 in two instances: Items 4 and 29. Further examination and discussion relating to Item 4 suggested an apparent cause of misfit. In this scenario, Laszlo (an older gentleman with memory troubles) is having a conversation with a couple that he thinks he knows, although it appears he is mistaken. Of the possible behavioral responses, Item 4 suggests “You leave Laszlo alone to finish his conversation.” When setting the scoring rubric, this was a very popular behavioral response consistent with person-centered care, because being too interventionist may undermine Laszlo. Thus, a correct score would require you to approve of this course of action. However, the precise wording “leave Laszlo alone” might have been interpreted as moving away and out of sight of Laszlo, or even abandoning him. Thus the wording was somewhat ambiguous and could have led to two differing interpretations by test-takers.

### Construct Validity

[Table 4](#) presents correlation coefficients for the SJT scores with five domains of personality. SJT scores were positively and statistically significantly correlated with “agreeableness” self-ratings ( $\rho = 0.25$ ). There was also a modest and statistically significant correlation between SJT scores and “open-mindedness” self-ratings. The other three “Big 5” personality traits (extraversion, conscientiousness, neuroticism) were not correlated with the SJT scores, as anticipated.

### Test–Retest Reliability

For a subgroup of care naïve participants, data were collected at a second survey, repeating both the SJTs and the care perception questions. The first analysis examined test–retest reliability. For 72 paired scores at two administrations of the SJTs, a strong correlation was found ( $\rho = 0.77$ ). This degree of correlation is regarded acceptable for measurement of performance using SJTs ([Catano et al., 2012](#)). An intraclass correlation coefficient was also estimated as 0.77.

### Care Perceptions

[Table 5](#) presents the degree of agreement with the five statements of care work perceptions, both before taking SJTs and at the retest. This sample was restricted to the “care naïve” sample of 72 who completed both surveys. The table illustrates a trend toward improved perceptions of care work

**Table 4.** Correlation Coefficients (*p* values) Between SJT Score and Personality Domains

Variable	SJT Score	Open-mindedness	Conscientiousness	Extraversion	Agreeableness
Openness to experience	0.179 (0.005)				
Conscientiousness	0.003 (0.957)	0.191 (0.002)			
Extraversion	-0.060 (0.341)	0.151 (0.016)	0.413 ( $<0.001$ )		
Agreeableness	0.250 ( $<0.001$ )	0.260 ( $<0.001$ )	0.562 ( $<0.001$ )	0.331 ( $<0.001$ )	
Negative emotionality	0.026 (0.678)	0.0678 (0.285)	-0.493 (0.001)	-0.532 ( $<0.001$ )	-0.374 ( $<0.001$ )

Note: SJT = Situational Judgement Test.

on all five statements. A crude “care perceptions scale” was compiled using all five statements, assigning values of 1–5 across the “strongly agree” to “strongly disagree” response options, and summing across each. Statements relating to care work being “emotionally rewarding,” being “proud to care” and involving “lots of decisions” were reverse-scored. Higher scores on the scale were indicative of more positive perceptions of care work. Before the SJT, participants had a mean score of 20.26 ( $SD = 2.18$ ) which had increased to 21.36 ( $SD = 1.97$ ) by the second administration. This difference was statistically significant, but modest in size (Cohen’s  $d = 0.528$ ;  $t = 3.78$ ,  $p < .001$ ).

### Exploratory Analysis

For the full sample, the mean SJT score was 23.7 ( $SD = 6.48$ ) and a range from 6 to 40, from a theoretically maximum range of zero to 51. The scores were approximately normally distributed (Shapiro–Wilk test:  $p = .507$ ). Table 6 illustrates that SJT scores did not vary substantially by gender, age, education level, or whether the participant was disabled. However, scores were markedly *higher* for two groups: those whose ethnicity was White, and those *without* prior or current experience as a care worker. With respect to the former, ethnicity appeared confounded by acutely low scores from the “booster” sample. Those who found the SJTs “difficult” to complete also scored more highly than those who found the exercise “easy” (post hoc test (Bonferroni):  $p = .001$ ).

### Discussion

This research aimed to examine the properties of a new assessment of person-centered values, for use in care worker recruitment. It is the first assessment of its kind in this sector, which, in England alone, employs more than 1.6 million people. The key findings are that the measure is broadly acceptable to test-takers, realistic in its portrayal of care work, promotes positive views of care work, and correlates significantly with important personality traits. However, SJTs are only as good as their quality appraisal. First and foremost, a SJT is a measurement instrument and any examination of quality must begin with its psychometric properties. In other sectors, SJTs have not always performed well against established benchmarks. In medicine, common difficulties with traditional SJTs include “uninterpretable” structures on factor analysis, relatively low reliability, and weak correlations with personality traits against predictions (Tiffin et al., 2020), and

construct validity described in some quarters as a “hot mess” (McDaniel et al., 2016).

The SJTs presented in this article appear to have encouraging measurement properties in this first sample of test-takers. Reliability statistics are above typical thresholds seen as conducive for measurement. The Cronbach alpha ( $=0.85$ ) and test–retest correlation ( $=0.77$ ) far exceed those seen in other sectors, with one meta-analysis of SJTs finding a pooled alpha coefficient of just 0.46 (Catano et al., 2012), compared with psychometric norms of 0.70 (Tiffin et al., 2020). Although the Person Separation Index ( $=1.52$ ) failed to reach the “ideal” threshold ( $>2.0$ ), it still exceeds the minimum value necessary to discern distinct strata (Fisher, 1992). There is evidence of unidimensionality from a CFA, which allows for a simple interpretation of a single summary score. The Rasch analysis hints at other trace factors, although this is common in SJTs (Tiffin et al., 2020) due to dependency introduced by multiple items drawn from a common “critical incident.”

The potential validity of the SJTs is supported by correlations with appropriate personality traits. Specifically, those scoring well on these SJTs tended to be agreeable people, a concept encompassing compassion, friendliness, gentleness, and trustworthiness (Crowe et al., 2018). Research has also indicated that more agreeable care workers are less likely to behave abusively (Chen et al., 2020). Care work is known to be a highly congruent career choice for those with a strong preference for interpersonal service jobs (Kuhn & Wolter, 2022), and these jobs tend to correlate highly with measured agreeableness (Barrick et al., 2003). A cautionary note is that the degree of association ( $\rho = 0.25$ ) is not large. The authors expected a “moderate” correlation coefficient but did not specify, *a priori*, what threshold to apply. In the wider literature, coefficients in this range are more commonly described as “weak-to-moderate” or even just “weak” (Abma et al., 2016).

A second noteworthy finding is the correlation between SJT scores with “openness to experience.” This implies that those with a more inventive or curious approach to care work will tend to score better on the SJTs than those who are more cautious and rules-driven. This is consistent with person-centered approaches to care, in which personalized and flexible support is strongly advocated over styles that are “too uniform; too analytical; too theoretical; too negative; and too directive” (Leplege et al., 2007; p. 1559). That SJTs were *not* correlated with other valuable psychological traits, in particular conscientiousness, further supports construct validity. Although it would certainly be possible to create SJTs that

**Table 5.** Agreement/Disagreement With Five Statements on the Perceptions of Care Work

Statement	T1	T2	Test statistic	
	<i>n</i> (%)	<i>n</i> (%)	$\chi^2$	Fisher's <i>p</i>
Care work is a low-skill job			(9) = 18.2	.025
Strongly agree	0 (0.0%)	0 (0.0%)		
Somewhat agree	3 (4.2%)	1 (1.4%)		
Neither agree or disagree	4 (5.6%)	2 (2.8%)		
Somewhat disagree	31 (43.1%)	22 (30.6%)		
Strongly disagree	34 (47.2%)	47 (65.3%)		
Care workers should be proud of their job			(6) = 15.3	.028
Strongly agree	61 (84.7%)	63 (87.5%)		
Somewhat agree	8 (11.1%)	8 (11.1%)		
Neither agree or disagree	2 (2.8%)	1 (1.4%)		
Somewhat disagree	1 (1.4%)	0 (0.0%)		
Strongly disagree	0 (0.0%)	0 (0.0%)		
Care work is emotionally rewarding			(12) = 49.9	<.001
Strongly agree	21 (29.2%)	25 (34.7%)		
Somewhat agree	41 (56.9%)	39 (54.2%)		
Neither agree or disagree	8 (9.5%)	7 (9.7%)		
Somewhat disagree	1 (1.2%)	1 (1.4%)		
Strongly disagree	1 (1.2%)	0 (0.0%)		
Care work involves making lots of decisions			(6) = 20.3	.046
Strongly agree	30 (41.7%)	37 (51.4%)		
Somewhat agree	33 (45.8%)	32 (44.4%)		
Neither agree or disagree	5 (6.9%)	3 (4.2%)		
Somewhat disagree	4 (5.6%)	0 (0.0%)		
Strongly disagree	0 (0.0%)	0 (0.0%)		
Care work is the same every day			(16) = 55.1	.002
Strongly agree	7 (9.7%)	4 (5.6%)		
Somewhat agree	28 (38.9%)	21 (29.2%)		
Neither agree or disagree	13 (18.1%)	9 (12.5%)		
Somewhat disagree	21 (29.2%)	32 (44.4%)		
Strongly disagree	3 (4.2%)	6 (8.3%)		

test conscientiousness, for example, it is not a key feature of person-centeredness and therefore not expected to correlate with these SJT scores.

A particularly encouraging finding is that the SJTs appear to be related to improved perceptions of care work. Long-term care, especially with older people, is known to face reputational challenges and stigma in attracting candidates to the sector (Manchha et al., 2022). Further, it is known that the recruitment process can signal important information to candidates about the occupation and organization, linked to their decision-making (Celani & Singh, 2011). SJTs have much in common with “Realistic Job Previews” in providing potential candidates with engaging, interactive scenarios that provide credible information on the likely fit between a person and the occupation (Klassen & Kim, 2021). Thus, showing potential applicants how challenging interpersonal situations likely to be encountered in the job role can be dealt with compassionately and effectively is likely to be attractive when recruiting to the sector.

However, of concern is the relatively lower scores observed for those identifying as of Black or Asian ethnicity. There is some evidence that SJTs in other fields tend to favor White

test-takers above ethnic minority groups, such as in medical education (Patterson et al., 2012). Some group differences may be linked to how people from different cultures interact with the SJT format, particularly “extreme response options” in Likert-type scales (Tiffin et al., 2020). Alternatively, there may have been a sampling bias attributable to the “booster” sample (see above), and potentially we could have designed the experiment with randomization of names/images to examine unconscious bias further. There may be some comfort taken from the fact that any bias toward White test-takers in SJT scores is certainly no larger than for other modes of candidate assessment (Lievens et al., 2016), and may even be smaller (Patterson et al., 2012). Further research is needed before firm conclusions can be reached.

The exploratory analysis found that incumbent care workers scored less well than the care naïve sample. This finding runs counter to many SJTs, because most operate as tests of “procedural knowledge” (Webster et al., 2020), and identifying the “right thing” to do in a given situation is something that is often learned on-the-job. However, as a tool intended to inspect the values-base of applicants, it is encouraging that incumbent care staff do not have an advantage on these tests.

**Table 6.** SJT Score by Socio-Demographic Characteristics

Characteristic	Mean	Standard deviation	Test statistic		
			<i>F</i>	<i>t</i>	<i>p</i>
Gender			0.68		.507
Female	24.0	6.22			
Male	22.8	7.66			
Other	23.0	2.83			
Age			0.62		.539
<30	24.2	6.50			
31-50	23.3	6.54			
51+	23.3	6.41			
Highest qualification			2.22		.086
Higher education	23.6	6.56			
A level or equivalent	25.5	6.12			
GCSE or equivalent	23.7	5.61			
None/below GCSE	21.5	7.55			
Ethnicity			11.41		<.001
Asian/Asian British	20.3	6.18			
Black/Black British	18.9	6.19			
White British	25.0	5.85			
White Other	26.9	5.39			
Mixed and other	22.9	8.82			
Disability				0.439	.669
Yes, has a disability	24.3	5.50			
No, not disabled	23.6	6.60			
Care worker				5.88	<.001
Incumbent care worker	21.8	6.45			
Care naïve	26.3	5.60			
Difficulty with SJTs			5.31		.002
Very easy	24.0	6.73			
Easy	22.5	6.13			
Neither easy nor difficult	23.4	6.69			
Difficult	27.2	5.64			
Very difficult	-	-			

Notes: GCSE = General Certificate of Secondary Education; SJT = Situational Judgement Test.

Explanations are most likely rooted in the use of people with lived experience as those who set the scoring rubric, rather than professionals who are typically used as “subject matter experts” in medical SJTs (Dunn et al., 2023). Care work in regulated domiciliary services in England has been criticized as providing safety-conscious and routinized support, which sits at odds with the wishes of older people (Backhouse & Ruston, 2022). Studies have highlighted that care workers with fixed “mindsets” in dementia care score less well on their knowledge of person-centered approaches (Kunz et al., 2022). The current study may thus reveal that workers have learned the rigid norms in their occupation, which are then reflected in their SJT responses. Arguably, these SJTs could serve to help change practice toward those behaviors and attitudes most desired by service users. However, this is a speculative interpretation, and further exploration is warranted.

### Strengths and Limitations

The study’s strengths are its relatively large sample amongst a mix of incumbent care workers and a care-naïve population.

This enables tests of both fidelity in the SJT representation of real-life care work but also acceptability and measurement properties among potential job-seekers. However, the findings need appraising in light of the study’s limitations. First, the samples (both incumbent care workers and care work “naïve”) were not randomly chosen, and there can be no claim that the responses are necessarily representative of either population. Scores achieved cannot be considered as a definitive benchmark against which future use of the tool can robustly be compared. A related limitation is that this is only the first study using these SJTs, and further corroborating evidence of measurement properties will generate more confidence. Second, it is important to recognize that SJTs are “low-fidelity” assessments, and we cannot make judgments about how well scores reflect actual behaviors in real-world situations. SJTs have been criticized as being susceptible to “faking,” or social desirability bias, whereby candidates understand and give the appropriate responses to SJT scenarios, but these are not the actions they would take in reality (Patterson, Zibarras, et al., 2016). It is also important to



recognize that these SJTs have only been tested in England, and their validity in international samples still needs to be established.

A final limitation is that no predictive validity of these SJTs has yet been conducted. SJTs in other sectors, including those with otherwise poor psychometric properties, rely on predictive validity as the essential “acid test” of whether the instrument is useful in recruitment and selection. In medicine, even with weak reliability and poor construct validity, the fact that they are found to predict job performance (measured variously, from supervisor ratings to complaints of clinical negligence) over-and-above academic credentials gives confidence in their usefulness (Patterson et al., 2016; Tiffin et al 2020). Without an assessment of whether the present SJTs correlate with job performance the evidence base for their validity will be weakened. A new study currently underway (at the time of writing) is making this assessment, together with a freely available platform to support their access.

### Implications for Policy and Practice

The many recruitment challenges facing the long-term care sector include the very real problem of finding people with the appropriate character. Care work is regarded as a values-based occupation, with quality bound by the virtues (or otherwise) of the person giving care. As found by Timonen and Lolich (2019) in their study with care managers and senior professionals: “participants spoke about the ‘proper caring person’ as someone who is essential for the entire process of care provision, yet remains elusive” (p. 735, emphasis in original). How, then, to make the elusive become tangible and tractable?

The translational impact of the situational judgement tests is now being realized in England and Wales. A new recruitment platform, [www.curiousaboutcare.org.uk](http://www.curiousaboutcare.org.uk), is freely available for all organizations recruiting personal care aides. These incorporate the full suite of SJTs, with a minor modification to Item 4 where Rasch’s analysis implied problematic wording (“leave Lazslo to finish his conversation” became “allow Lazslo ...”) No alteration to the second misfitting item could be so easily conducted, and this was retained in its existing form. The platform introduces a simple visualized scoring system, alongside a guide to implementation. The platform has been adopted by “Skills for Care,” the nationwide government-sponsored sector skills body, which is supporting launch and local adoption. Further research is now underway to assess implementation lessons and supporting behavioral science to examine how the platform impacts decision-making processes.

### Conclusions

Situational Judgment Tests are used in a wide range of industries and occupations to evaluate important but non-technical skills and knowledge. These are central to care work, particularly with older people, where interpersonal abilities and judgments are crucial to the user experience. This paper finds that a set of SJTs designed for care work have encouraging properties in this first sample of test-takers. Specifically, the SJTs are accepted by incumbent care workers as representing the reality of the job; they appear correlated with improved before-and-after assessments of the social standing of care work; and they have positive psychometric properties.

Furthermore, the SJTs correlate as expected with key personality traits. However, further work is needed to examine SJT implementation and to establish whether SJT scores correlate with job performance.

### Supplementary Material

Supplementary data are available at *Innovation in Aging* online.

### Funding

This report is independent work funded by the Abbeyfield Research Foundation and the National Institute for Health and Care Research School for Social Care Research (Ref: P143). The views expressed in this publication are those of the author(s) and not necessarily those of the Abbeyfield Research Foundation, NIHR SSCr, the National Institute for Health and Care Research or the Department of Health and Social Care.

### Conflict of Interest

None.

### Data Availability

Anonymous data sought for replication purposes are available on reasonable request, subject to ethical permissions.

### Acknowledgments

We acknowledge the support of many long-term care organizations and care workers that supported the study.

### References

- Abma, I. L., Rovers, M., & van der Wees, P. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. *BMC Research Notes*, 9, 226. <https://doi.org/10.1186/s13104-016-2034-2>
- Abrams, R., Vandrevalla, T., Samsi, K., & Manthorpe, J. (2019). The need for flexibility when negotiating professional boundaries in the context of home care, dementia and end of life. *Ageing & Society*, 39(9), 1976–1995. <https://doi.org/10.1017/s0144686x18000375>
- Antwi, Y. A., & Bowblis, J. R. (2018). The impact of nurse turnover on quality of care and mortality in nursing homes: Evidence from the Great Recession. *American Journal of Health Economics*, 4(2), 131–163. [https://doi.org/10.1162/ajhe\\_a\\_00096](https://doi.org/10.1162/ajhe_a_00096)
- Backhouse, T., & Ruston, A. (2022). Home-care workers’ experiences of assisting people with dementia with their personal care: A qualitative interview study. *Health & Social Care in the Community*, 30(3), e749–e759. <https://doi.org/10.1111/hsc.13445>
- Barrick, M. R., Dustin, S. L., Giluk, T. L., Stewart, G. L., Shaffer, J. A., & Swider, B. W. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology*, 85(2), 330–352. <https://doi.org/10.1111/j.2044-8325.2011.02036.x>
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland’s occupational types. *Personnel Psychology*, 56(1), 45–74. <https://doi.org/10.1111/j.1744-6570.2003.tb00143.x>

- Catano, V. M., Brochu, A., & Lamerson, C. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, 20(3), 333–346. <https://doi.org/10.1111/j.1468-2389.2012.00604.x>
- Celani, A., & Singh, P. (2011). Signaling theory and applicant attraction outcomes. *Personnel Review*, 40(2), 222–238. <https://doi.org/10.1108/00483481111106093>
- Chen, W., Fang, F., Chen, Y., Wang, J., Gao, Y., & Xiao, J. (2020). The relationship between personality traits, caring characteristics and abuse tendency among professional caregivers of older people with dementia in long-term care facilities. *Journal of Clinical Nursing*, 29(17–18), 3425–3434. <https://doi.org/10.1111/jocn.15380>
- CIPD. (2015). *A head for hiring: The behavioural science of recruitment and selection*. Chartered Institute of Personnel and Development, London.
- Crowe, M. L., Lynam, D. R., & Miller, J. D. (2018). Uncovering the structure of agreeableness from self-report measures. *Journal of Personality*, 86(5), 771–787. <https://doi.org/10.1111/jopy.12358>
- D'Astous, V., Abrams, R., Vandrevale, T., Samsi, K., & Manthorpe, J. (2019). Gaps in understanding the experiences of homecare workers providing care for people with dementia up to the end of life: A systematic review. *Dementia (London, England)*, 18(3), 970–989. <https://doi.org/10.1177/1471301217699354>
- Davison, I., McManus, C., & Taylor, C. (2016). *Evaluation of GP specialty selection*. <https://research.birmingham.ac.uk/en/publications/evaluation-of-gp-specialty-selection>
- Delp, L., Wallace, S. P., Geiger-Brown, J., & Muntaner, C. (2010). Job stress and job satisfaction: Home care workers in a consumer-directed model of care. *Health Services Research*, 45(4), 922–940. <https://doi.org/10.1111/j.1475-6773.2010.01112.x>
- Dunn, A., Tiffin, P. A., & Wilberforce, M. (2023). Promoting person-centred care in the home support of older people through situational judgement tests. *International Journal of Care and Caring*, 1–16. <https://doi.org/10.1332/239788221x16887228610432>
- Fisher, W. P. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions*, 6(3), 238.
- Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2), 765–800. <https://doi.org/10.1093/qje/qjx042>
- Hussein, S. (2017). “We don’t do it for the money” ... The scale and reasons of poverty-pay among frontline long-term care workers in England. *Health & Social Care in the Community*, 25(6), 1817–1826. <https://doi.org/10.1111/hsc.12455>
- Klassen, R. M., & Kim, L. E. (2021). Beyond selection: Applying lessons from teacher selection to recruiting and developing teacher candidates. In R. M. Klassen & L. E. Kim (Eds.), *Teacher selection: Evidence-based practices* (pp. 147–158). Springer International Publishing. [https://doi.org/10.1007/978-3-030-76188-2\\_10](https://doi.org/10.1007/978-3-030-76188-2_10)
- Kuhn, A., & Wolter, S. C. (2022). Things versus people: Gender differences in vocational interests and in occupational preferences. *Journal of Economic Behavior & Organization*, 203, 210–234. <https://doi.org/10.1016/j.jebo.2022.09.003>
- Kunz, L. K., Scheibe, S., Wisse, B., Boerner, K., & Zemlin, C. (2022). From dementia mindsets to emotions and behaviors: Predicting person-centered care in care professionals. *Dementia (London, England)*, 21(5), 1618–1635. <https://doi.org/10.1177/14713012221083392>
- Lepage, A., Gzil, F., Cammelli, M., Lefevre, C., Pachoud, B., & Ville, I. (2007). Person-centredness: Conceptual and historical perspectives. *Disability and Rehabilitation*, 29(20–21), 1555–1565. <https://doi.org/10.1080/09638280701618661>
- Leverton, M., Burton, A., Beresford-Dent, J., Rapaport, P., Manthorpe, J., Mansour, H., Guerra Ceballos, S., Downs, M., Samus, Q., Dow, B., Lord, K., & Cooper, C. (2021). “You can”t just put somebody in a situation with no armour”: An ethnographic exploration of the training and support needs of homecare workers caring for people living with dementia. *Dementia*, 20(8), 2982–3005. <https://doi.org/10.1177/14713012211023676>
- Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: Evidence from the UKCAT. *Medical Education*, 50(6), 624–636. <https://doi.org/10.1111/medu.13060>
- Linacre, J. M. (2003). Rasch power analysis: Size vs significance: Standardized chi-square fit statistic. *Rasch Measurement Transactions*, 17(1), 918.
- Llena-Nozal, A., Rocard, E., & Sillitti, P. (2022). Providing long-term care: Options for a better workforce. *International Social Security Review*, 75(3–4), 121–144. <https://doi.org/10.1111/issr.12310>
- Manchha, A. V., Way, K. A., Tann, K., & Thai, M. (2022). The social construction of stigma in aged-care work: Implications for health professionals’ work intentions. *The Gerontologist*, 62(7), 994–1005. <https://doi.org/10.1093/geront/gnac002>
- Manthorpe, J., & Bramley, S. (2020). Barriers and facilitators to ex-service personnel obtaining employment within social care roles with older people: A scoping review. *Working with Older People*, 24(3), 205–220. <https://doi.org/10.1108/wwop-04-2020-0011>
- Manthorpe, J., Harris, J., Samsi, K., & Moriarty, J. (2017). Doing, being and becoming a valued care worker: User and family carer views. *Ethics and Social Welfare*, 11(1), 79–91. <https://doi.org/10.1080/17496535.2016.1247904>
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63–91. <https://doi.org/10.1111/j.1744-6570.2007.00065.x>
- McDaniel, M. A., List, S. K., & Kepes, S. (2016). The “hot mess” of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology*, 9(1), 47–51. <https://doi.org/10.1017/iop.2015.115>
- Ministry of Justice. (2021). *Recruitment, assessment, and retention in the direct care workforce for individuals with criminal records: A comprehensive model approach*. US Department of Justice, Washington.
- Morgan, J. C., Dill, J., & Kalleberg, A. L. (2013). The quality of health-care jobs: Can intrinsic rewards compensate for low extrinsic rewards? *Work Employment and Society*, 27(5), 802–822. <https://doi.org/10.1177/0950017012474707>
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *The Journal of Applied Psychology*, 91(4), 749–761. <https://doi.org/10.1037/0021-9010.91.4.749>
- Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook* (pp. 97–146). Springer US. [https://doi.org/10.1007/978-1-4684-2490-4\\_4](https://doi.org/10.1007/978-1-4684-2490-4_4)
- OECD. (2020). *Who cares? Attracting and retaining care workers for the elderly*. OECD Publishing, Paris. <https://doi.org/10.1787/92c0ef68-en>
- Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, 46(9), 850–868. <https://doi.org/10.1111/j.1365-2923.2012.04336.x>
- Patterson, F., Prescott-Clements, L., Zibarras, L., Edwards, H., Kerrin, M., & Cousans, F. (2016). Recruiting for values in healthcare: A preliminary review of the evidence. *Advances in Health Sciences Education: Theory and Practice*, 21, 859–881. <https://doi.org/10.1007/s10459-014-9579-4>
- Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No 100. *Medical Teacher*, 38(1), 3–17. <https://doi.org/10.3109/0142159X.2015.1072619>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin*, 28(6), 789–801. <https://doi.org/10.1177/0146167202289008>

- Scales, K. (2021). It is time to resolve the direct care workforce crisis in long-term care. *The Gerontologist*, 61(4), 497–504. <https://doi.org/10.1093/geront/gnaa116>
- Skills for Care. (2022). *The state of the adult social care sector and workforce in England*. Skills for Care, London.
- Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69–81. <https://doi.org/10.1016/j.jrp.2017.02.004>
- Spetz, J., Stone, R. I., Chapman, S. A., & Bryant, N. (2019). Home and community-based workforce for patients with serious illness requires support to meet growing needs. *Health affairs (Project Hope)*, 38(6), 902–909. <https://doi.org/10.1377/hlthaff.2019.00021>
- Tiffin, P. A., Paton, L. W., O'Mara, D., MacCann, C., Lang, J. W. B., & Lievens, F. (2020). Situational judgement tests for selection: Traditional vs construct-driven approaches. *Medical Education*, 54(2), 105–115. <https://doi.org/10.1111/medu.14011>
- Timonen, V., & Lolich, L. (2019). “The poor carer”: Ambivalent social construction of the home care worker in elder care services. *Journal of Gerontological Social Work*, 62(7), 728–748. <https://doi.org/10.1080/01634372.2019.1640334>
- Webster, E. S., Paton, L. W., Crampton, P. E. S., & Tiffin, P. A. (2020). Situational judgement test validity for selection: A systematic review and meta-analysis. *Medical Education*, 54(10), 888–902. <https://doi.org/10.1111/medu.14201>
- Wilberforce, M., Challis, D., Davies, L., Kelly, M. P., Roberts, C., & Clarkson, P. (2017). Person-centredness in the community care of older people: A literature-based concept synthesis. *International Journal of Social Welfare*, 26(1), 86–98. <https://doi.org/10.1111/ijsw.12221>
- Zallman, L., Finnegan, K. E., Himmelstein, D. U., Touw, S., & Woolhandler, S. (2019). Care for America's elderly and disabled people relies on immigrant labor. *Health Affairs (Project Hope)*, 38(6), 919–926. <https://doi.org/10.1377/hlthaff.2018.05514>