the british psychological society
promoting excellence in psychology

British Journal of Health Psychology (2020), 25, 474–501

www.wileyonlinelibrary.com

Acceptability, reliability, and validity of a brief measure of capabilities, opportunities, and motivations ("COM-B")

Chris Keyworth[1]*, Tracy Epton[1], Joanna Goldthorpe[1], Rachel Calam[1] and Christopher J. Armitage[1,2]

[1]Division of Psychology and Mental Health, Faculty of Biology, Medicine and Health, Manchester Centre for Health Psychology, School of Health Sciences, The University of Manchester, UK

[2]Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, UK

Objectives. The Capabilities, Opportunities, Motivations, Behaviour (COM-B) model is being used extensively to inform intervention design, but there is no standard measure with which to test the predictive validity of COM or to assess the impact of interventions on COM. We describe the development, reliability, validity, and acceptability of a generic 6-item self-evaluation COM questionnaire.

Design and methods. The questionnaire was formulated by behaviour change experts. Acceptability was tested in two independent samples of health care professionals (N = 13 and N = 85, respectively) and a sample of people with low socio-economic status (N = 214). Acceptability (missing data analyses and user feedback), reliability (test–retest reliability and Bland–Altman plots) and validity (floor and ceiling effects, Pearson's correlation coefficient [r], exploratory factor analysis [EFA], and confirmatory factor analysis [CFA]) were tested using a national survey of 1,387 health care professionals.

Results. The questionnaire demonstrated acceptability (missing data for individual items: 5.9–7.7% at baseline and 18.1–32.5% at follow-up), reliability (ICCs .554–.833), and validity (floor effects 0.6–5.5% and ceiling effects 4.1–22.9%; pairwise correlations rs significantly <1.0). The regression models accounted for between 21 and 47% of the variance in behaviour. CFA (three-factor model) demonstrated a good model fit, ($\chi^2[6] = 7.34$, $p = .29$, RMSEA = .02, CFI = .99, TLI = .99, BIC = 13,510.420, AIC = 13,428.067).

Conclusions. The novel six-item questionnaire shows evidence of acceptability, validity, and reliability for self-evaluating capabilities, opportunities, and motivations. Future research should aim to use this tool in different populations to obtain further support for its reliability and validity.

**Correspondence should be addressed to Dr Chris Keyworth, Division of Psychology and Mental Health, Faculty of Biology, Medicine and Health, Manchester Centre for Health Psychology, School of Health Sciences, The University of Manchester, Coupland 1 Building – Room G3, Oxford Road, Manchester M13 9PL, UK (email: chris.keyworth@manchester.ac.uk).*

DOI:10.1111/bjhp.12417

Statement of contribution

What is already known on the subject?

- The Capability, Opportunity, Motivation (COM), Behaviour (-B) model is being used extensively to inform intervention design.
- The lack of an accepted universal measure hinders progress in behaviour change.

What does this study add?

- There is evidence of acceptability, validity, and reliability for self-evaluating COM.
- Our measure may be sufficiently generic for any behaviour or population, although this requires further testing.

The UK National Institute for Health and Care Excellence cites the *Capabilities, Opportunities, Motivations, Behaviour (COM-B*) model (Michie, van Stralen, & West, 2011) as a key theoretical framework for understanding and supporting behaviour change (National Institute for Health & Care Excellence, 2014). The COM-B model comprises six components that are hypothesized to drive behaviour, namely physical capability (having the skills, strength, and stamina); psychological capability (being able to engage in the necessary thought processes such as comprehension and reasoning); physical opportunity (afforded by the environment, including time and resources); social opportunity (afforded by interpersonal influences, social cues, and how we think about things, such as the words and concepts that make up language); reflective motivation (conscious intentions, plans, and making evaluations); and automatic motivation (emotional reactions, impulses, and desires; Michie, Atkins, & West, 2014; Michie, van Stralen, & West, 2011).

COM-B is designed to provide an overarching model that captures all the factors known to influence behaviour change (Michie *et al.*, 2014). Its origins can be found in health behaviour models such as the theory of planned behaviour, health belief model, social cognitive theory, protection motivation theory, self-determination theory, transtheoretical model, and the health action process approach (Michie, van Stralen, *et al.*, 2011). Meta-analyses show that these models can explain up to 37% of the variance in behaviour, which is 'large' in Cohen's (1992) terms, for example, the theory of planned behaviour (Armitage & Conner, 2001), the theory of reasoned action (McEachan *et al.*, 2016; Sheppard, Hartwick, & Warshaw, 1988), the health belief model (Harrison, Mullen, & Green, 1992), social cognitive theory (Young, Plotnikoff, Collins, Callister, & Morgan, 2014), self-determination theory (Plotnikoff, Costigan, Karunamuni, & Lubans, 2013), and the transtheoretical model (Plotnikoff *et al.*, 2013). As an overarching model, one would anticipate that the COM-B model will similarly explain large proportions of the variance in behaviour, and may even exceed the predictive validity of rivalling models of health behaviour.

Use of the COM-B model is widespread with respect to: (a) guiding data collection and analysis in qualitative studies (Atkins, Kelly, Littleford, Leng, & Michie, 2017), (b) informing intervention development (Barker, Atkins, & de Lusignan, 2016), and (c) explaining the findings of systematic reviews (Simon & West, 2015). However, despite the widespread use of COM-B, there is currently no standard measure that operationalizes fully each of the six domains of the model. This is important because it means there is a lack of evidence as to the predictive validity of the model and there is no tool with which to

evaluate the mechanisms of action with respect to the impact of COM-based interventions on behaviour. The aim of the present research was to develop and test such a measure.

Current COM-B questionnaires

Seven COM-B questionnaires have been reported in the literature to date, but it is not clear how they have been developed and mapped onto the corresponding six components of the COM-B model and their associated definitions (Ayton *et al.*, 2017; Balku *et al.*, 2017; Hankonen *et al.*, 2017; Howlett, Schulz, Trivedi, Troop, & Chater, 2017; Stevely *et al.*, 2018; Taylor *et al.*, 2016; Webb, Hall, Hall, & Fabunmi-Alade, 2016). Moreover, just two of the previous questionnaires have been assessed for reliability or validity: Howlett et al. found evidence of psychological capability and reflective motivation as predictors of physical activity, and Ayton et al. found evidence of adequate internal consistency and construct validity in measuring perceived exercise capability, opportunity, and motivation. In these two studies, however, the absence of measures of acceptability may limit conclusions about the perceived appropriateness and relevance of the questionnaire, and the lack of test–retest analyses limits conclusions about the reliability of the items (Howlett *et al.*, 2017; Taylor *et al.*, 2016). Additionally, there is a need to develop a brief measure of COM. Whilst questionnaires typically range from between 10 (Stevely *et al.*, 2018) and 19 (Taylor *et al.*, 2016) items, the longest of the existing questionnaires comprised 194 items (Balku *et al.*, 2017).

Consequently, there is a need to develop the first generic self-evaluation questionnaire to assess people's perceptions of capabilities, opportunities, and motivations, intended for use in multiple behaviours and a range of diverse populations, including patients, health care professionals, and general population samples, particularly given that previous questionnaires have been developed for use in specific contexts. Addressing limitations of previous studies, the present research aimed to develop a new questionnaire to operationalize fully the six subdomains of the COM-B model reliably and validly using rigorous psychometric evaluation.

Aims

The aims of this study were to describe the development and evaluate the reliability, validity, and acceptability of a measure to assess perceived *capabilities*, *opportunities*, and *motivations* in relation to behaviour change. The primary aim of this study was to develop a measure that may be sufficiently generic to enable adaptation and testing across behaviours and populations.

Methods

Overview

Ethical approval was obtained from a university research ethics committee (ref: 2017-0739-1780), and informed consent was obtained from participants. We followed STROBE reporting guidelines (presented in Supplementary File A). Our COM measure was developed in three phases. In phase 1, the items were developed through consensus using the expertise within the research team, and distributed to a small convenience sample of health care professionals for initial feedback. A second round of more in-depth piloting was conducted with a new sample of health care professionals recruited using study

Table 1. Two measures of acceptability and number (%) of participants expressing positive, negative, and neutral comments/feedback for each question

	Participant rating		Comment type			
Question	Mean	(*SD*)	Positive	Negative	Neutral	No comment
Physical opportunity			5 (2%)	8 (4%)	36 (17%)	165 (77%)
Ease of reading	8.10	2.41				
Understanding	8.04	2.27				
Social opportunity			0 (0%)	7 (3%)	26 (12%)	181 (85%)
Ease of reading	8.24	2.28				
Understanding	8.09	2.26				
Reflective motivation			3 (1%)	0 (0%)	35 (16%)	176 (82%)
Ease of reading	8.92	1.54				
Understanding	8.76	1.67				
Automatic motivation			4 (2%)	0 (0%)	22 (10%)	188 (88%)
Ease of reading	8.89	1.54				
Understanding	8.65	1.75				
Physical capability			3 (1%)	0 (0%)	38 (18%)	173 (81%)
Ease of reading	8.85	1.59				
Understanding	8.71	1.72				
Psychological capability			2 (1%)	2 (0%)	24 (11%)	186 (87%)
Ease of reading	8.77	1.65				
sUnderstanding	8.53	1.85				

advertisements through health care professional bodies. To assess generalizability, the questionnaire was then piloted among a general population sample with low socio-economic status and a new target behaviour (phase 2). For the main study (phase 3), a national sample of health care professionals were recruited via a survey panel company (YouGov), as part of a larger study examining the prevalence of health care professionals delivering opportunistic behaviour change interventions (Keyworth *et al.*, 2018).

Phase 1: Researcher development and initial piloting

COM-B instrument design and development. To address the identified gaps in the literature for a COM instrument, the researchers designed a six-item questionnaire to assess *physical capability, psychological capability, physical opportunity, social opportunity, reflective motivation,* and *automatic motivation*.[1]

Phase 2 (pilot study with low-SES sample)

In order to ensure that the questionnaire was generic and suitable for as broad an audience as possible, we tested the questionnaire with a sample of people with low socio-economic status (SES) using an alternative target behaviour (behaviour change to improve health; the questionnaire is presented in full in Supplementary File C).[2]

[1] *The design and development process is outlined in full in Supplementary File B, and the questionnaire is presented in Supplementary File C.*

[2] *The full process is outlined in Supplementary File B.*

Table 1 shows that the questionnaire was rated favourably in terms of ease of reading and understanding. Low numbers of participants reported dissatisfaction with the items, with three of six items receiving no negative comments (reported in Table 1). Open-text comments were then coded by a member of the research team into one of three categories: positive, negative, or neutral (whereby comments were neither positive nor negative). Comments were often related to participants' own health affecting the likelihood of doing more exercise (e.g., limitations due to physical disability or existing health problem), or more general beliefs about the factors involved in behaviour change. Therefore, the research team agreed that the results of phase 2 did not warrant any changes to the questionnaire.

Phase 3 (main study)

Health care professionals with a patient-facing role were recruited via a survey panel company (YouGov), as part of a larger study examining the prevalence of health care professionals delivering opportunistic behaviour change interventions (Keyworth *et al.*, 2018). A purposive sample of health care professionals working in the National Health Service (NHS) in the United Kingdom were invited to take part in an online questionnaire and were incentivized in accordance with YouGov's points system (respondents accumulate points for taking part in surveys, which can be exchanged for cash or entry into a prize draw).

A total of 1,387 health care professionals completed the questionnaire at Time 1 and included nurses and health visitors (N = 438), GPs (N = 332), scientific, therapeutic and technical staff (N = 270), and specialist doctors (N = 125). Participants were mostly female (N = 941; 67.8%), with a mean age of 45 years. Demographics are presented in full elsewhere (Keyworth *et al.*, 2018). All participants who completed the questionnaire at Time 1 (N = 1,387) were invited to take part in a 1-month follow-up study, and were told this was to 'test the statistical reliability of the questionnaire'. A recruitment flow diagram is presented in Figure 1, displaying the number of participants taking part in each phase of the study and reasons for non-participation. Participants who consented to take part in the 1-month follow-up were asked to provide their email addresses to be contacted again. A total of 426 health care professionals were sent an email invitation to complete the Time 2 questionnaire; 209 health care professionals completed the 1-month follow-up. A total of 43 participants did not supply an email address, and we were therefore unable to match the follow-up data with the baseline data. A total of 166 health care professionals took part in the 1-month follow-up questionnaire. The chi-square was used to gauge the representativeness of the follow-up sample (n = 166) compared with the baseline (n = 1,387) sample. The baseline and follow-up samples were comparable in age, ethnicity, and seven of the nine health care professional categories (X^2 values are presented in Table 2).

Procedure

Six items measuring *capabilities, opportunities,* and *motivations* were included as part of a cross-sectional survey distributed online to a sample of health care professionals working in the NHS with a patient-facing role. Data were collated by YouGov and sent securely to the research team for analysis. A member of the research team sent invitations

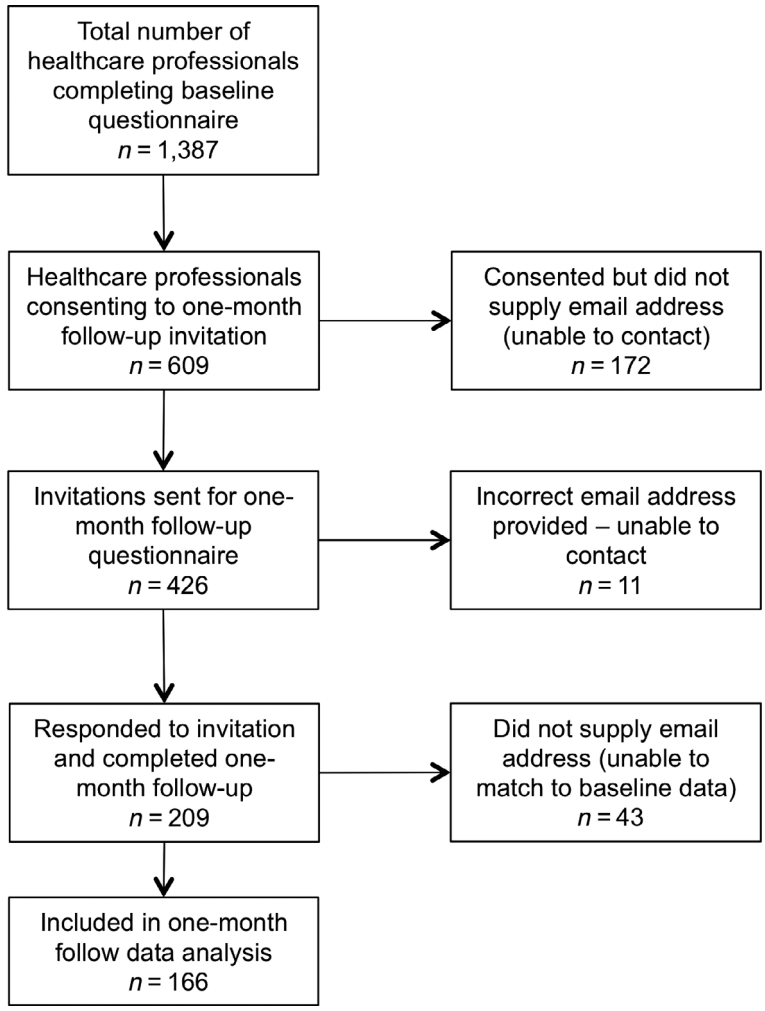

Figure 1. Recruitment flow diagram for 1-month follow-up data collection.

for the Time 2 questionnaire to consenting participants 1 month after completion of the Time 1 questionnaire.

Analysis

Data were analysed using descriptive and inferential statistics to examine the acceptability, reliability, and validity of the six items. Acceptability was assessed using: (1) missing data analyses; (2) three quality indicators to assess perceived ease of understanding, interest, and balance and fairness of the questionnaire; and (3) a content analysis of open-ended comment boxes included in the questionnaire. Reliability was assessed using: (1) test–retest reliability (intra-class correlation coefficients) and (2) Bland–Altman plots. Content validity was assessed by examining floor and ceiling effects. Discriminant validity was assessed using interitem correlations (Pearson's *r*). Predictive validity was assessed using multiple regression analyses. A flow diagram showing the steps involved at each phase is presented in Figure 2.

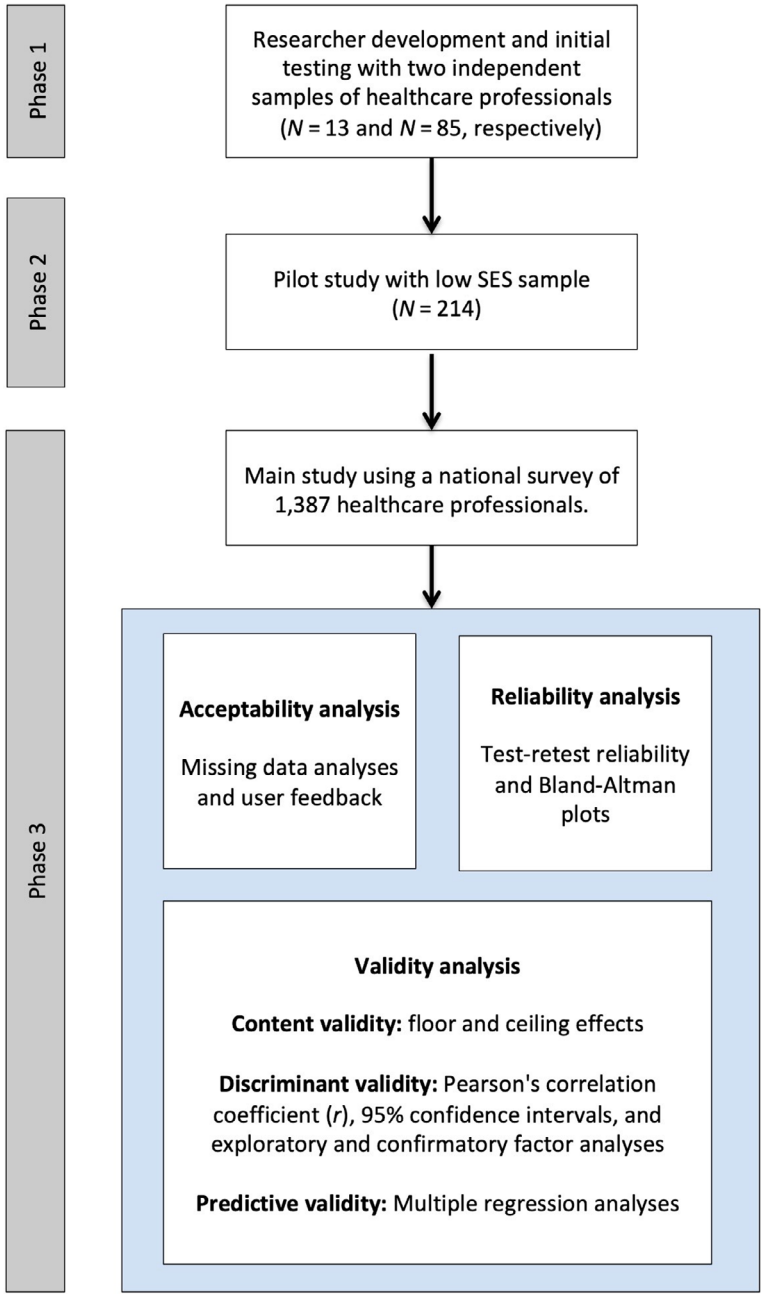

Figure 2. Flow diagram of questionnaire development process. [Colour figure can be viewed at wileyonlinelibrary.com]

Acceptability analysis

Missing data analyses

Descriptive statistics are presented for: (1) the number of participants completing all six items and (2) the number of missing responses for each item. Data are presented as descriptive statistics.

Quality indicators

Three items to assess the quality of the questionnaire were included at the end of the survey[3]: (1) 'Overall, how easy or difficult did you find it to understand the questions?' (on a rating scale from: [1] 'difficult to understand' to [9] 'easy to understand'); (2) 'Overall, how boring or interesting did you find the questionnaire?' (on a rating scale from: [1] 'boring' to [9] 'interesting'); and (3) 'Overall, how fair and balanced did you find the questions?' (on a rating scale from: [1] 'none of the questions were fair and balanced' to [9] 'all of the questions were fair and balanced'). A series of one-sample *t*-tests were conducted to assess how far above or below the neutral rating (5) people rated the four quality indicator scales. This measure was used to assess desirability and acceptability of the questionnaire.

Content analysis of open-ended text

Participants were invited to provide open-ended comments describing any difficulties completing the questionnaire, along with any more general feedback including asking participants to describe any aspects of the questionnaire that were unclear. Two questions were asked: (1) 'Do you have any comments on your experience of taking this survey?'; and (2) 'Do you have any other comments, such as whether the questions made assumptions about respondents, didn't display correctly on common screen sizes and formats, could have caused offence, had any grammatical/ spelling errors, or other things?' User comments were combined, and a content analysis was performed.

Reliability analysis

Test–retest reliability

Intra-class correlation coefficients (ICCs) were used to determine test–retest reliability. A series of two-way mixed-effects models with measures of absolute agreement were used. ICCs were determined as <.40 (poor), .40–.75 (fair to good), and >.75 (excellent; Fleiss, 1986).

Bland–Altman plots

Bland–Altman plots (Bland & Altman, 1999) were used to examine any systematic differences between test and retest scores, and therefore establish agreement between the two scores. This was calculated using the mean differences between the two scores, standard deviations, and limits of agreement (mean of the differences $\pm$ $1.96 \times SD$).

Validity analysis

Content validity

Floor and ceiling effects were calculated for each of the six items at Time 1 and Time 2. Data are presented according to the number of respondents who answered each item at both the minimum and maximum points of each item (e.g., 0 and 10, or 0 and 100). Floor and ceiling effects can be used to determine content validity (Terwee *et al.*, 2007;

[3] *The three items are included as quality indicators as part of every YouGov survey.*

Wamper, Sierevelt, Poolman, Bhandari, & Haverkamp, 2010). High floor and ceiling effects may lead to difficulties in: (1) distinguishing participants from each other; and (2) measuring changes in participants' ratings before and after an intervention (Terwee *et al.*, 2007; Wamper *et al.*, 2010). The recognized value of 15% of the sample is used to determine whether floor and ceiling effects are observed with the proportion of responses being at either the minimum or maximum point of the items (McHorney & Tarlov, 1995).

Discriminant validity

Two methods were used to establish discriminant validity. First, Pearson's correlation coefficient (r) was used to assess the strength of the relationship between the items. Pearson's r is interpreted as .10 (small effect), .30 (moderate effect), and .50 (large effect) (Cohen, 1988). A series of pairwise correlations were conducted to examine relationships between the six items. As each item is deemed to measure a different construct (*physical capability, psychological capability, physical opportunity, social opportunity, reflective motivation,* and *automatic motivation*), low correlation between items overall was expected (Pearson's $r < 0.50$). Items within each subdomain of COM were expected to be more highly correlated (pair 1: *physical capability* and *psychological capability;* pair 2: *physical opportunity* and *social opportunity;* and pair 3: *reflective motivation* and *automatic motivation*) with an expected Pearson's $r > 0.50$.

Second, following the recommendations of Kline (2011) discrimination can also be established if the upper limit of the 95% confidence interval of each pairwise correlation does not exceed the suggested threshold of .85. Thus, we expected the least discriminating cases to occur between the related subdomains.

To explore further whether the items within the questionnaire are related, exploratory factor analysis (EFA) was used to assess the factor structure within the six items. A correlation matrix was created to assess the relationships between the items. Bartlett's test of sphericity and the Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) were assessed to determine the suitability of factor analysis for the data set. KMO should be a minimum of 0.6 (Tabachnick *et al.*, 2007). For items to show correlation, Bartlett's test of sphericity should be $p < 0.05$ (Field, 2009). As we expected correlations between the components of COM, EFA was performed with three fixed factors, using direct oblimin rotation (to permit correlations between factors), with principal axis factoring. Factors were considered as salient if they were greater than .40 (Stevens, 1992).

Follow-up CFA ($n = 373$) and EFA ($n = 373$). To explore the factor structure further, we used a random number generator to divide the sample into two equal groups to conduct an additional EFA, and to conduct a confirmatory factor analysis (CFA), using STATA version 14, to verify the fit of a 3-factor model (373 sets of responses for both the CFA and EFA). For the CFA, we used maximum-likelihood estimation and evaluated model fit following the recommendations of Kline (2011): the chi-square statistic, the comparative fit index (CFI), the Tucker–Lewis Index (TLI), and the root mean square error of approximation (RMSEA). Based on Kline (2011), a good model fit is evidenced by a non-significant chi-square, a CFI and TLI of at least .95, and an RMSEA of .05 or less. Standardized factor loadings were expected to be >.4 (Giesinger *et al.*, 2016).

Table 2. Comparison of baseline (n = 1,387) and follow-up (n = 166) samples used for test–retest reliability

Variable	Baseline				Follow-up				X^2 for difference between baseline and follow-up X^2 for difference between baseline and follow-up
	n	%	Mean	(*SD*)	*n*	%	Mean	(*SD*)	
Gender									
Male	446	(32.2)			37	(22.3)			21.38 ($p < .001$)
Female	941	(67.8)			129	(77.7)			21.38 ($p < .001$)
Total	1,387								
Age (years)			45 .00	(11.46)			45.11	(11.30)	
Ethnicity									
White or White British	1,209	(87.2)			153	(92.2)			1.33 ($p = .25$)
Other	177	(12.8)			13	(7.8)			1.33 ($p = .25$)
Did not state	1	(0.1)							
Total	1,387				166				
Health care professional group									
General practitioners	332	(23.9)			3	(1.8)			21.40 ($p < .001$)
Specialist doctors	125	(9.0)			20	(12.0)			0.48 ($p = .49$)
Nurses and health visitors	438	(31.6)			65	(39.2)			1.07 ($p = .30$)
Midwives	42	(3.0)			5	(3.0)			0 ($p = 1.00$)
Ambulance staff	20	(1.4)			5	(3.0)			1.02 ($p = .31$)
Scientific, therapeutic, and technical staff	270	(19.5)			48	(28.9)			2.19 ($p = .14$)
Nurses working in GP practices	88	(6.3)			0	(0)			4.30 ($p < .05$)
Support to clinical staff	49	(3.5)			8	(4.8)			0.12 ($p = .73$)
Other HCHS staff/unknown classifications	23	(1.3)			0	(0)			1.01 ($p = .32$)
Total	1,387				166				

Predictive validity

A series of multiple regression models were used to examine independent association between the COM variables with: (1) delivery of opportunistic behaviour change interventions; and (2) time spent delivering opportunistic behaviour change interventions. A separate regression model was used for each COM variable for each dependent variable, and each model was adjusted for potential confounders (age, gender, ethnicity, and profession).

Results

Acceptability analysis

Missing data analyses

The results of the missing data analyses at Time 1 and Time 2 are presented in Table 3. At Time 1, there were a total of 1,387 respondents, with all items having fewer than 10% missing data points (range = 5.9–7.7%). At 1-month follow-up, 166 participants were included in the analysis. Missing data for each item were higher than observed at Time 1 (range = 18.1–32.5%). There were missing data on all six items for thirty participants.

Completion rates for each item are presented in Table 3. A total of 1,181 participants (85% of the total sample) completed all six measures at Time 1; 102 participants (61% of the sample included in the 1-month follow-up analysis) completed all measures (answering on the rating scale of either 0–100 or 0–10, or by answering 'don't know'). At Time 1, completion rates for each item (excluding the 'don't know' option) ranged from 70 to 82.1%. Items most likely to be completed were *automatic motivation* (82.1% of participants answered), psychological capability (81.9% of participants answered), and physical capability (81.1% of participants answered). The item least likely to be completed was social opportunity (70% of participants answered). The number of 'don't know' responses at Time 1 ranged from 10.2 to 22.5% across all items. The highest number of 'don't know' responses was observed for the items *physical opportunity* (n = 288, 20.7%) and *social opportunity* (n = 312, 22.5%). At Time 2, completion rates for each item ranged from 63.9 to 78.9%. Items most likely to be completed were *reflective motivation* (78.9% of participants answered), and *physical capability* and *psychological capability* (for both items, 78.3% of participants answered). Mirroring the Time 1 findings, the item least likely to be completed was *social opportunity* (63.9% of participants answered). The number of 'don't know' responses was lower than observed at Time 1; responses were less than 4% (range = 2.4–3.6%).

The proportion of participants scoring at each point on the rating scale of each item is presented in Table 4. There are two key findings. First, there were a high proportion of responses at the lower end of the *physical opportunity* and *social opportunity* items, compared with the other items, at Time 1 and Time 2 (21% and 35% of participants respectively, rating 0–10 on the 0–100 scales). Second, there were a high proportion of responses at the upper end of the *physical capability* and *psychological capability* items, compared with the other items, at Time 1 and Time 2 (21% and 29%, and 22% and 24% of participants respectively, rating 10 on the 0–10 scales).

Table 3. Missing data and 'don't know' responses at Time 1 (n = 1,387) and Time 2 (n = 166)

	Time 1							Time 2						
		Complete responses[1]		Don't know		Did not answer			Complete responses[1]		Don't know		Did not answer	
Item	Total	*n*	%	*n*	%	*n*	%	Total	*n*	%	*n*	%	*n*	%
Physical opportunity	1,387	1,015	73.2	288	20.7	84	6.0	166	111	66.9	4	2.4	51	30.7
Social opportunity	1,387	970	70.0	312	22.5	105	7.6	166	106	63.9	6	3.6	54	32.5
Reflective motivation	1,387	1,112	80.2	193	13.9	82	5.9	166	131	78.9	4	2.4	31	18.7
Automatic motivation	1,387	1,139	82.1	141	10.2	107	7.7	166	128	77.1	6	3.6	32	19.3
Physical capability	1,387	1,125	81.1	157	11.3	105	7.6	166	130	78.3	6	3.6	30	18.1
Psychological capability	1,387	1,136	81.9	153	11.0	98	7.1	166	130	78.3	4	2.4	32	19.3

[1]Calculated according to: total number of participants (Time 1 n = 1,387; Time 2 n = 166) − ('Don't know' responses + 'did not answer' responses).

Table 4. Proportion of responses at each point according to item at Time 1 and Time 2

Item	Responses at each point (*n*; %)											Total
PO[1]	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100		
Time 1	218 (21)	92 (9)	140 (14)	43 (4)	106 (10)	82 (8)	45 (4)	101 (10)	25 (2)	163 (16)		1,016
Time 2	39 (35)	20 (18)	6 (5)	2 (2)	13 (12)	2 (2)	1 (1)	6 (5)	4 (4)	18 (16)		111
SO[2]	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100		
Time 1	269 (28)	76 (8)	110 (11)	48 (5)	98 (10)	97 (10)	48 (5)	91 (9)	35 (4)	98 (10)		971
Time 2	26 (25)	17 (16)	10 (9)	3 (3)	14 (13)	7 (7)	2 (2)	8 (8)	7 (7)	12 (11)		106
RM[3]	0	1	2	3	4	5	6	7	8	9	10	
Time 1	54 (5)	40 (4)	48 (4)	34 (3)	48 (4)	142 (13)	123 (11)	176 (16)	163 (15)	113 (10)	171 (15)	1,113
Time 2	2 (2)	2 (2)	6 (5)	5 (4)	5 (4)	14 (11)	11 (8)	25 (19)	30 (23)	8 (6)	23 (18)	131
AM[4]	0	1	2	3	4	5	6	7	8	9	10	
Time 1	77 (7)	44 (4)	81 (7)	45 (4)	30 (3)	116 (10)	92 (8)	159 (14)	202 (18)	109 (10)	184 (16)	1,140
Time 2	7 (5)	5 (4)	6 (5)	5 (4)	0 (0)	13 (10)	13 (10)	18 (14)	30 (23)	10 (8)	21 (16)	128
PhC[5]	0	1	2	3	4	5	6	7	8	9	10	
Time 1	64 (1)	50 (4)	61 (5)	49 (4)	26 (2)	91 (8)	82 (7)	156 (14)	187 (17)	123 (11)	236 (21)	1,126
Time 2	1 (1)	1 (1)	3 (3)	4 (3)	2 (2)	15 (12)	6 (5)	19 (15)	30 (23)	11 (8)	38 (29)	130
PsC[6]	0	1	2	3	4	5	6	7	8	9	10	
Time 1	43 (4)	40 (4)	50 (4)	32 (3)	31 (3)	98 (9)	85 (7)	151 (13)	216 (19)	140 (12)	250 (22)	1,137
Time 2	3 (2)	1 (1)	3 (2)	3 (2)	2 (2)	7 (5)	11 (8)	22 (17)	30 (23)	17 (13)	31 (24)	130

[1]Physical opportunity; [2]social opportunity; [3]reflective motivation; [4]automatic motivation; [5]physical capability; [6]psychological capability.

Quality indicators

Participants were asked to rate the quality of the questionnaire based on three indicators: 'understanding', 'interest', and 'balance and fairness'. One-sample *t*-tests showed participants rated the understanding score, $t(492) = 31.56$, $p \leq .001$, the interest score, $t(499) = 34.36$, $p \leq .001$, and balance and fairness score, $t(495) = 36.26$, $p \leq .001$ significantly higher than the scale midpoints, demonstrating acceptability. For the 'understanding' indicator ($M = 7.47$, $SD = 1.74$), most participants rated this as either 8 ($n = 105$) or 9 ($n = 188$). For the 'interest' indicator ($M = 7.55$, $SD = 1.66$), participants rated this as either 8 ($n = 110$) or 9 ($n = 190$). For the 'balance and fairness' indicator ($M = 7.56$, $SD = 1.57$), most participants rated this as either 8 ($n = 121$) or 9 ($n = 179$). For the overall rating ($M = 7.36$, $SD = 1.74$), most participants rated this as either 8 ($n = 234$) or 9 ($n = 325$).

Content analysis of open-ended comment boxes

Fifty-three participants provided open-text comments in answer to the questionnaire feedback items. A content analysis was performed with responses coded into the most prominent categories.

Seven participants expressed dissatisfaction with the questionnaire items. Specific items were not reported as ambiguous, but participants provided general comments relating to clarity of the wording of the questionnaire items. The most prominent category of comments was 'technical difficulties, formatting and layout'. Illustrative quotes are provided in Table 5. Sixteen participants (30.2% of the sample who provided comments) described technical difficulties navigating through the questionnaire. Five (9.4% of the sample who provided comments) health care professionals stated that they did not see the topic of making every content count as relevant to their role.

Reliability analysis

Test–retest reliability

Results are presented in Table 6. Data are analysed according to participants who completed each item at baseline and follow-up: *physical opportunity* ($n = 95$), *social opportunity* ($n = 94$), *reflective motivation* ($n = 123$), *automatic motivation* ($n = 123$), *physical capability* ($n = 122$), and *psychological capability* ($n = 127$). Test–retest reliability was fair to good for four of the six items (ICC .554–.707): *Physical opportunity, social opportunity, physical capability,* and *psychological capability*. Two items were rated as excellent (ICC > .75*): reflective motivation* (ICC .830) and *automatic motivation* (ICC .833).

Bland–Altman plots

Bland–Altman plots are presented in Supplementary File D (figures 1–6; heat maps are also presented in figures 7–12), with the mean of the test and retest scores presented on the *x*-axis and the difference between the two scores presented on the *y*-axis. Bias (i.e., mean difference) and 95% levels of agreement were 7.84 (−67.94, 83.62) for *physical opportunity*, 2.50 (−58.94, 63.88) for *social opportunity*, −0.30 (−4.10, 3.50) for *reflective motivation*, −0.08 (−4.48, 4.65) for *automatic motivation*, −0.71 (−5.27, 4.65) for *physical capability*, and −0.42 (−5.27, 4.43) for *psychological capability*.

Table 5. Content analysis categories relating to the two open-ended questions

Category	Description	Illustrative quote
Technical difficulties, formatting and layout	Participants expressed technical difficulties navigating through the questionnaire. These included slow loading times or pages failing to display correctly.	'Experienced a few frustrating technical problems'. (Dietician) 'Had technical problems and had to abandon survey but was able to pick up again where I left off' (Nurse) 'I wanted to go back to review some of my answers and was unable to' (Nurse)
Questions perceived as irrelevant	Some participants perceived the questions as being irrelevant to their health care professional role	'Quite a few questions were not relevant to me even though they were for a clinician as not all clinicians are involved in interacting with the "service users" in the same way' (Anaesthetist) 'About the right length but I am not sure that I am qualified to answer the type of questions that were posed' (Pathologist)
Positive feedback		'Best survey, most relevant I have done' (Nurse) 'Nice to be asked' (Cardiologist)

Table 6. Reliability demonstrated by intra-class correlation coefficient (ICC) and 95% confidence intervals (CI) for COM items

	Reliability data	
Item	ICC	95% CI
Physical opportunity ($n = 95$)[a]	.554**	0.336–0.701
Social opportunity ($n = 94$)[b]	.707**	0.561–0.805
Reflective motivation ($n = 123$)[c]	.830**	0.758–0.881
Automatic motivation ($n = 123$)[d]	.833**	0.761–0.883
Physical capability ($n = 122$)[e]	.608**	0.438–0.727
Psychological capability ($n = 127$)[f]	.674**	0.538–0.770

Notes. Excludes 'don't know' responses.
[a]Excludes participants who answered 'don't know' at both Time 1 and Time 2 ($n = 2$) and participants who switched between a 'don't know' and a valid response across the time points ($n = 17$); [b]Excludes participants who switched between a 'don't know' and a valid response across the time points ($n = 11$). No participants answered 'don't know' at both Time 1 and Time 2; [c]Excludes participants who answered 'don't know' at both Time 1 and Time 2 ($n = 2$) and participants who switched between a 'don't know' and a valid response across the time points ($n = 8$); [d]Excludes participants who answered 'don't know' at both Time 1 and Time 2 ($n = 1$) and participants who switched between a 'don't know' and a valid response across the time points ($n = 7$); [e]Excludes participants who answered 'don't know' at both Time 1 and Time 2 ($n = 3$) and participants who switched between a 'don't know' and a valid response across the time points ($n = 9$); [f]Excludes participants who answered 'don't know' at both Time 1 and Time 2 ($n = 2$) and participants who switched between a 'don't know' and a valid response across the time points ($n = 5$).; **$p < .001$.

Limits of agreement were wide, and there was evidence of heteroscedasticity of the points around the null line (an expected finding when using bounded scales). Further analysis for the *physical capability* item resulted in a significant bias of −0.71 (2.73) ($p < .01$), and thus, agreement between test and retest scores could not be established. Agreement was established for all other items.

Validity analysis

Content validity

Floor and ceiling effects were calculated for each of the six items assessing COM at both Time 1 and Time 2. Data are presented according to those who answered each item at both the minimum and maximum points of each item (e.g., 0 and 10, or 0 and 100). Time 1 results are presented in Table 7. No ceiling effects were observed for *physical opportunity, social opportunity, reflective motivation,* and *automatic motivation.* However, ceiling effects were observed for *physical capability* and *psychological capability*. One-month follow-up results are presented in Table 7. Results mirrored the Time 1 findings, with ceiling effects found for the same two items.

Discriminant validity

Pearson's *r* is interpreted as .10 (small effect), .30 (medium effect), and .50 (large effect) (Cohen, 1988). A series of pairwise correlations are presented. Time 1 results are reported in Table 8. All correlations were medium to large (r = .429–.783). The two *capability* items (r = .698), the two *opportunity* items (r = .783), and the two *motivation* items (r = .762) were the most highly intercorrelated. In contrast, at Time 2 (Table 8), seven correlations were small (r = .159–.297), and eight correlations were medium to large (r = .370–.738). The most highly correlated items, which mirrored the findings of the Time 1 results, were between the two *capability* items (r = .616), the two *opportunity* items (r = .608), and the two *motivation* items (r = .738).

Recommended thresholds provided by Kline (2011) were used to establish discriminant validity. At Time 1 (Table 8), the upper limit of the 95% confidence interval of each pairwise correlation did not exceed the suggested threshold of 0.85, thus suggesting evidence of discriminant validity. The least discriminating cases were between the pairs of items expected to be the most highly correlated; *physical opportunity* and *social opportunity*, *reflective motivation* and *automatic motivation*, and *physical capability* and *psychological capability*. For *physical opportunity* and *social opportunity*, the upper confidence interval limit was .807; for *reflective motivation* and *automatic motivation*, the upper confidence interval limit was .786; and for *physical capability* and

Table 7. Floor and ceiling effects (Time 1 and Time 2)

	Time 1					Time 2				
		Floor effect		Ceiling effect			Floor effect		Ceiling effect	
Item	Total responses	*N*	%	*N*	%	Total responses	*N*	%	*N*	%
Physical opportunity	1,387	50	3.6	93	6.7	166	5	3	15	9.0
Social opportunity	1,387	62	4.5	57	4.1	166	5	3	11	6.6
Reflective motivation	1,387	54	3.9	170	12.3	166	2	1.2	23	13.9
Automatic motivation	1,387	77	5.5	183	13.2	166	7	4.2	21	12.7
Physical capability	1,387	64	4.6	235	16.9[1]	166	1	0.6	38	22.9[1]
Psychological capability	1,387	43	3.1	249	18.0[1]	166	3	1.8	31	18.7[1]

[1]Floor/ceiling effect observed, where 15% of data is observed at the minimum or maximum points of each item.

Table 8. Pearson's correlations in relation to Time 1[a] and Time 2[b] data

Item	Physical opportunity	Social opportunity	Reflective motivation	Automatic motivation	Physical capability	Psychological capability	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Physical opportunity	–	.608[3] (.465–.720) (.06)	.180[1] (−.008 to .356) (.09)	.186[1] (−.004 to .363) (.09)	.383[2] (.209–.533) (.09)	.289[1] (.106–.453) (.09)	44.79 (33.56)	39.11 (35.45)
Social opportunity	.783[3] (.756–.807) (.02)	–	.159[1] (−.035 to .341) (.09)	.175[1] (−.019 to .356) (.09)	.405[2] (.231–.554) (.09)	.387[2] (.212–.538) (.09)	40.20 (32.57)	42.64 (33.06)
Reflective motivation	.526[3] (.486–.564) (.03)	.572[3] (.526–.615) (.02)	–	.738[3] (.646–.809) (.06)	.251[1] (.080–.407) (.09)	.370[2] (.210–.511) (.08)	6.42 (2.80)	6.90 (2.45)
Automatic motivation	.488[2] (.438–.535) (.03)	.544[3] (.496–.588) (.03)	.762[3] (.736–.786) (.02)	–	.297[1] (.129–.449) (.09)	.381[2] (.210–.511) (.08)	6.27 (3.05)	6.57 (2.88)
Physical capability	.442[2] (.389–.492) (.03)	.451[2] (.398–.501) (.03)	.532[3] (.487–.574) (.03)	.542[3] (.498–.583) (.03)	–	.616[3] (.495–.713) (.07)	6.60 (3.06)	7.62 (2.29)
Psychological capability	.429[2] (.375–.480) (.03)	.458[2] (.405–.508) (.03)	.533[3] (.488–.575) (.03)	.572[3] (.531–.611) (.02)	.698[3] (.666–.736) (.02)	–	6.96 (2.83)	7.55 (2.33)

Notes. Numbers in parentheses are standard errors.

AM = automatic motivation; PhC = physical capability; PsC = psychological capability; PO = physical opportunity; RM = reflective motivation; SO = social opportunity

Correlations below the diagonal column refer to Time 1; correlations above the diagonal column refer to Time 2

[1]Small relationship; [2]Medium relationship; [3]Large relationship; [a]Pairwise correlations based on the following sampling sizes: PO/SO, $n = 908$; PO/RM, $n = 923$; PO/AM, $n = 938$; PO/PhC, $n = 942$; PO/PsC, $n = 943$; SO/RM, $n = 890$; SO/AM, $n = 903$; SO/PhC, $n = 908$; SO/PsC, $n = 908$; RM/AM, $n = 1{,}065$; RM/PhC, $n = 1{,}054$; RM/PsC, $n = 1{,}058$; AM/PhC, $n = 1{,}077$; AM/PsC, $n = 1{,}084$; and PhC/PsC, $n = 1{,}082$. Numbers in parentheses are standard errors; [b]Pairwise correlations based on the following sampling sizes: PO/SO, $n = 97$; PO/RM, $n = 109$; PO/AM, $n = 107$; PO/PhC, $n = 108$; PO/PsC, $n = 109$; SO/RM, $n = 104$; SO/AM, $n = 103$; SO/PhC, $n = 105$; SO/PsC, $n = 106$; RM/AM, $n = 125$; RM/PhC, $n = 127$; RM/PsC, $n = 128$; AM/PhC, $n = 126$; AM/PsC, $n = 126$; PhC/PsC, $n = 128$.

psychological capability, the upper confidence interval limit was .807. Time 2 results (Table 8) mirrored the Time 1 results; all upper confidence interval limits did not exceed the suggested threshold of 0.85. The least discriminating cases were between the pairs of items expected to be the most highly correlated, consistent with the Time 1 findings. Thus, we observe strong evidence for discriminant validity among the six COM items.

To explore further the relationship between the six items, exploratory factor analysis (EFA) was conducted to assess the underlying factor structure. EFA was performed with three fixed factors, using direct oblimin rotation. The KMO value was .809, suggesting our sampling was adequate to conduct factor analysis (Tabachnick *et al.*, 2007), and Bartlett's test of sphericity was significant ($p < .001$), suggesting both an adequate sample size (Field, 2009), and that items were sufficiently correlated to run the analysis.

Items loaded onto three factors (item loadings are presented in Table 9, and interfactor correlations are presented in Table 10), which explained 86% of the total variance. As expected, and mirroring the results of the Pearson's *r* correlation analysis, items were loaded onto three factors that corresponded to *capability, opportunity,* and *motivation* items (pair 1: *physical capability* and *psychological capability;* pair 2: *physical opportunity* and *social opportunity;* and pair 3: *reflective motivation* and *automatic motivation*). However, whilst the three factors explained 57.4%, 9.7%, and 5.3% of the variance, respectively, only one of the corresponding eignenvalues was greater than one (3.72, .87, and .59, respectively). Contrary to the theoretical assumptions of the COM model, this may suggest a unidimensional solution according to Kaiser criterion (eigenvalues > 1). We therefore conducted further analyses.

We used the Bayesian information criterion (BIC) and Akaike information criterion (AIC) to compare the three-factor model with the unidimensional solution. The model with the lowest values of BIC and AIC is to be preferred (Raftery, 1995).

Table 9. Exploratory factor analysis loadings for the six items ($N = 746$)

	Factors[a]		
Item	1	2	3
Psychological capability	***.901***	−.428	−.560
Physical capability	***.898***	−.452	−.553
Physical opportunity	.449	***−.944***	−.553
Social opportunity	.471	***−.938***	−.591
Reflective motivation	.560	−.579	***−.947***
Automatic motivation	.611	−.566	***−.942***

[a]Highest factor loading in bold.

Table 10. Interfactor correlations from the exploratory factor analysis

	Factor 1	Factor 2	Factor 3
Factor 1	1	−.59	−.75
Factor 2	−.59	1	.69
Factor 3	−.75	.69	1

Note. Interfactor correlations demonstrate a high correlation between factor 1 and factor 2 (−.59), between factor 1 and factor 3 (−.75), and between factor 2 and factor 3 (.69)

*Follow-up CFA (*n = *373) and EFA (*n = *373).* Confirmatory factor analysis demonstrated a good fit for the three-factor model, $\chi^2(6) = 7.34$, $p = .29$, RMSEA = .02, CFI = .99, TLI = .99, BIC = 13510.420, AIC = 13428.067 (see Figure 3, top diagram). All parameters in the model were significant at $p < .001$, and all standardized factor loadings exceeded the recommended .40 threshold. The bivariate correlations between *capability* and *opportunity* ($r = .58$), *capability* and *motivation* ($r = .73$), and *opportunity* and *motivation* ($r = .69$) were all significant at $p < .001$. In contrast, the unidimensional CFA model exhibited poor model fit, $\chi^2(9) = 210.12$, $p < .001$, RMSEA = .25, CFI = .84, TLI = .73, BIC = 13695.434, AIC = 13624.846. However, all parameters in the model were significant at $p < .001$, and all standardized factor loadings exceeded the recommended .40 threshold (see Figure 3, bottom diagram). When comparing the models, the BIC and AIC favoured a three-factor solution, as opposed to a unidimensional solution (indicated by lower BIC and AIC values when comparing the two models).

Using the same criteria as the earlier EFA (KMO = .797 and Bartlett's test of sphericity was $p < .001$), items were loaded onto three factors and explained 74% of the total variance. The three factors explained 57.5, 10.6, and 5.9% of the variance, respectively, with only one of the corresponding eignenvalues greater than 1 (3.71, .90, and, .62).

Predictive validity (descriptive statistics)

For both of our principal outcomes, namely the proportion of patients with whom health care professionals deliver opportunistic behaviour change interventions, and the proportion of the consultation time spent on delivering interventions, participants rated their physical opportunity ($M = 44.96$, $SD = 33.13$; $M = 43.75$, $SD = 33.42$) and social opportunity ($M = 40.58$, $SD = 32.45$; $M = 39.61$, $SD = 32.75$) as being statistically significantly lower than the midpoint of the scales (Table 11). For the remaining items, the means ranged from 6.35 to 7.01 for delivery of interventions, and 6.27 to 6.91 for time spent delivering interventions.

The zero-order correlations showed that, for both outcomes, the COM domains physical opportunity, social opportunity, reflective motivation, automatic motivation, physical capability, and psychological capability were significantly and positively correlated with delivery of behaviour change interventions. Correlations ranged from $r = .36$ to $r = .60$ (mean $r = .52$) in the case of delivery of interventions (Table 11), and $r = .31$ to $r = .53$ (mean $r = .45$) in the case of time spent on delivering interventions.

Predictive validity (Predicting delivery of opportunistic behaviour change interventions)

Table 12 shows the results of the multiple regression analyses used to test the association of the components of the COM domains with the proportion of patients with whom health care professionals deliver opportunistic behaviour change interventions.

After adjustment for the potential cofounders (age, gender, ethnicity, and profession), all of the COM variables in the analyses resulted in a statistically significant R^2, accounting for between 21% and 39% of the variance observed (all ps $< .001$). The standardized beta weights showed that *physical opportunity* ($\beta = 0.49$, $p < .001$), *social opportunity* ($\beta = 0.51, p < .001$), *reflective motivation* ($\beta = 0.54, p < .001$), *automatic motivation* ($\beta = 0.54$, $p < .001$), *physical capability* ($\beta = 0.36$, $p < .001$), and *psychological capability* ($\beta = 0.28, p < .001$) were all significant predictors of health care professionals

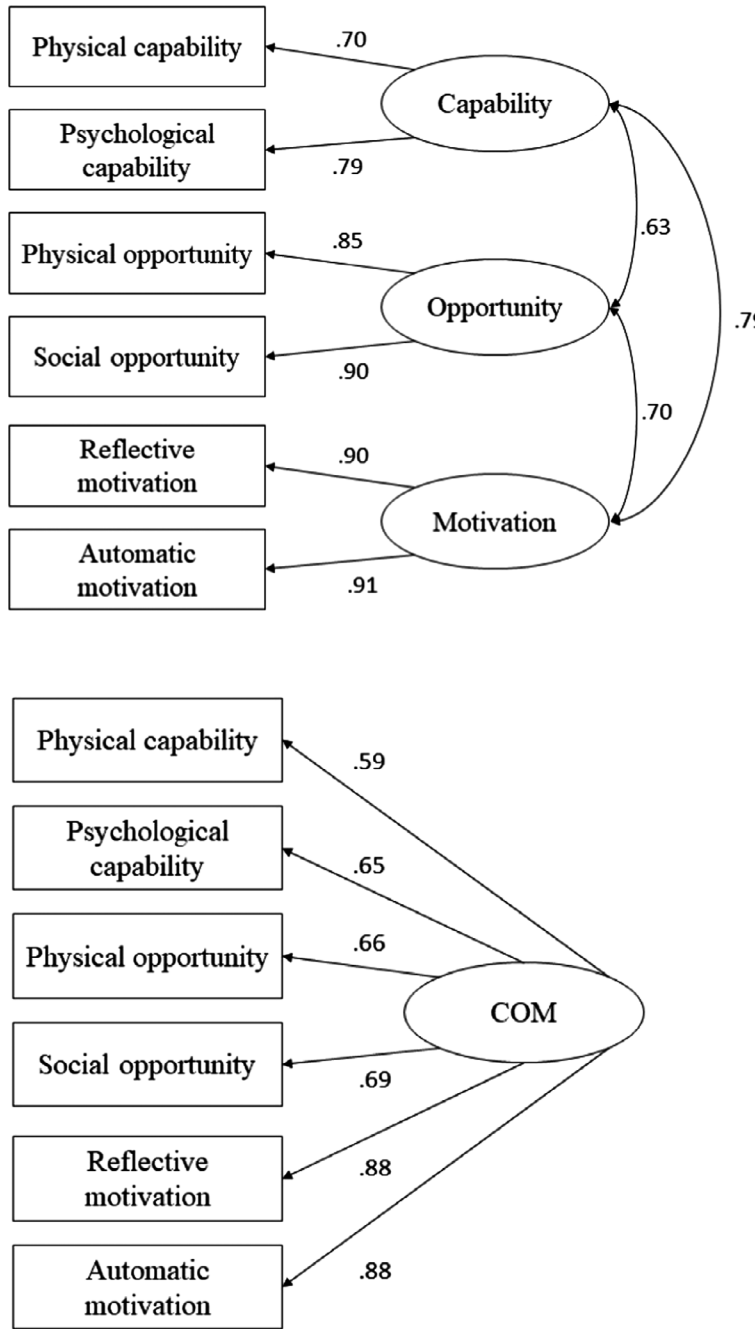

Figure 3. Confirmatory factor analysis of the capability, opportunity, and motivation questionnaire (n = 373). All parameters are significant at $p < .001$, for both the three-factor solution (top) and the unidimensional solution.

delivering opportunistic behaviour change interventions. Therefore, the results demonstrated that all six of the COM domains are strong predictors of health care professional practice.

Table 11. Descriptive statistics and zero-order correlations: *delivery* of opportunistic behaviour change interventions and *time spent* delivering interventions

	1.	2.	3.	4.	5.	6.	7.	Frequency *M* (*SD*)	Time *M* (*SD*)
Physical opportunity	–	.78*** (0.03)	.55*** (0.03)	.50*** (0.03)	.42*** (0.03)	.43*** (0.03)	.55***	44.96 (33.13)	43.75 (33.42)
Social opportunity	.78*** (0.02)	–	.60*** (0.03)	.55*** (0.03)	.43*** (0.03)	.44*** (0.03)	.57***	40.58 (32.45)	39.61 (32.75)
Reflective motivation	.56*** (0.03)	.59*** (0.03)	–	.79*** (0.02)	.55*** (0.03)	.54*** (0.03)	.59***	6.50 (2.84)	6.40 (2.85)
Automatic motivation	.50*** (0.03)	.55*** (0.03)	.80*** (0.02)	–	.54*** (0.03)	.57*** (0.03)	.60***	6.35 (3.03)	6.27 (3.05)
Physical capability	.41*** (0.03)	.43*** (0.03)	.55*** (0.03)	.54*** (0.03)	–	.68*** (0.03)	.43***	6.59 (3.09)	6.49 (3.12)
Psychological capability	.42*** (0.03)	.44*** (0.03)	.53*** (0.03)	.56*** (0.02)	.68*** (0.02)	–	.36***	7.01 (2.84)	6.91 (2.88)
Delivery/time spent (proportion)	.49***	.51***	.53***	.48***	.36***	.31***	–	50.00 (31.15)	35.30 (30.92)

Correlations above the diagonal column refer to delivery of interventions; correlations below the diagonal column refer to time spent delivering interventions *$p < .05$; **$p < .01$; and ***$p < .001$.; *Pairwise correlations based on the following sampling sizes, where PO = physical opportunity, SO = social opportunity, RM = reflective motivation, AM = automatic motivation, PhC = physical capability, PsC = psychological capability: (1) delivery of interventions: PO/SO, $n = 796$; PO/RM, $n = 800$; PO/AM, $n = 810$; PO/PhC, $n = 809$; PO/PsC, $n = 810$; SO/RM, $n = 791$; SO/AM, $n = 797$; SO/PhC, $n = 801$; SO/PsC, $n = 798$; RM/AM, $n = 885$; RM/PhC, $n = 877$; RM/PsC, $n = 880$; AM/PhC, $n = 892$; AM/PsC, $n = 896$; and PhC/PsC, $n = 892$, and (2) time spent delivering interventions: PO/SO, $n = 770$; PO/RM, $n = 770$; PO/AM, $n = 776$; PO/PhC, $n = 779$; PO/PsC, $n = 776$; SO/RM, $n = 759$; SO/AM, $n = 763$; SO/PhC, $n = 771$; SO/PsC, $n = 765$; RM/AM, $n = 835$; RM/PhC, $n = 834$; RM/PsC, $n = 833$; AM/PhC, $n = 843$; AM/PsC, $n = 842$; and PhC/PsC, $n = 845$. Numbers in parentheses are standard errors.

Table 12. Multiple regression analysis predicting health care professionals *delivering* opportunistic behaviour change interventions

Predictor	R^{2a}	β^a	Adjusted R^{2a}	Adjusted β^a
Physical opportunity	.31	.55***	.36	.49***
Social opportunity	.32	.57***	.36	.51***
Reflective motivation	.35	.59***	.37	.54***
Automatic motivation	.36	.60***	.39	.54***
Physical capability	.18	.43***	.24	.36***
Psychological capability	.13	.36***	.21	.28***

Notes. Models were run separately for each component of 'COM'.
[a]Standardized *$p \leq .05$, **$p \leq .01$, and ***$p \leq .001$; [b]Model adjusted for potential confounders: age, gender, ethnicity, and profession.

Predictive validity (predicting time spent on opportunistic behaviour change interventions)
Table 13 shows the results of the multiple regression analyses used to test the association of the components of the COM domains with time spent delivering opportunistic behaviour change interventions.

After adjustment for the potential cofounders (age, gender, ethnicity, and profession), all of the COM variables in the analyses resulted in a statistically significant R^2, accounting for between 24% and 47% of the variance observed (all ps $< .001$). The standardized beta weights showed that *physical opportunity* ($\beta = 0.42$, $p < .001$), *social opportunity* ($\beta = 0.44, p < .001$), *reflective motivation* ($\beta = 0.47, p < .001$), *automatic motivation* ($\beta = 0.41$, $p < .001$), *physical capability* ($\beta = 0.29$, $p < .001$), and *psychological capability* ($\beta = 0.24$, $p < .001$) were all significant predictors of time spent delivering opportunistic behaviour change interventions. Therefore, the results demonstrated that all six of the COM domains are strong predictors of health care professional practice.

Discussion

This paper describes the development of the first brief, generic measure of a 6-item self-evaluation questionnaire designed to assess perceived *capabilities* (physical and psychological), *opportunities* (physical and social), and *motivations* (reflective and automatic). This is the first study to focus on the development and testing of the psychometric properties of a brief questionnaire based on the COM-B model (Michie *et al.*,

Table 13. Multiple regression analysis predicting health care professionals *time spent* delivering opportunistic behaviour change interventions

Predictor	R^{2a}	β^a	Adjusted R^{2a}	Adjusted β^a
Physical opportunity	.24	.49***	.31	.42***
Social opportunity	.26	.51***	.32	.44***
Reflective motivation	.28	.53***	.33	.47***
Automatic motivation	.23	.48***	.29	.41***
Physical capability	.13	.36***	.22	.29***
Psychological capability	.10	.31***	.20	.24***

Notes. Models were run separately for each component of 'COM'.
[a]Standardized *$p \leq .05$, **$p \leq .01$, and ***$p \leq .001$; [b]Model adjusted for potential confounders: age, gender, ethnicity, and profession.

2014) that fully operationalizes all six COM subdomains, developed in accordance with the recognized definitions of the components of the COM-B model (Michie *et al.*, 2014), and intended for use across study populations, including patients, health care professionals, and general population samples. There was evidence of good acceptability of the questionnaire, and evidence of test–retest reliability, and discriminant and predictive validity.

There are four key findings. First, test–retest reliability was fair to good for four of the six items included in the questionnaire (*physical opportunity, social opportunity, physical capability,* and *psychological capability*) and excellent for two items (*reflective motivation* and *automatic motivation*). Whilst this demonstrates stability in item ratings over time, results must be interpreted with caution given the intra-class correlation coefficients were based on participants who provided a response on the rating scale at both time points only, and do not account for participants who answered 'don't know' at either time point.

Second, there is evidence for discriminant validity. Effect size for the interitem correlations ranged from small to large. The items most highly correlated at Time 1 and Time 2 were the two *capability* items, the two *opportunity* items, and the two *motivation* items. Results of our exploratory factor analysis showed that items were loaded onto three factors and corresponded to *capability, opportunity,* and *motivation* items. However, there was evidence of discriminant validity at Time 1 and Time 2; the upper limit of the 95% confidence interval of each pairwise correlation did not exceed the recommended threshold (Kline, 2011). Results therefore suggest that the subitems in each COM domain showed some relationship, but the items across domains were deemed to be measuring different constructs, as is expected in the literature (Michie, van Stralen, et al., 2011).

Third, there is evidence supporting the use of the COM model of behaviour for predicting health care professionals' delivery of opportunistic behaviour change interventions. This is the first time the predictive validity of the COM-B model has been tested in this context. The COM-B variables accounted for large proportions of the variance observed in self-reported behaviour (delivering opportunistic behaviour change interventions and time spent delivering interventions; R^2 = .47 and .35, respectively). Results from meta-analyses of rivalling behaviour change models show that they account for up to 37% of the variance in behaviour. In the present study, capability, opportunity, and motivation accounted for 47% of the variance in delivery of opportunistic behaviour change interventions and 35% of the variance in time spent delivering interventions. Thus, COM-B explains as much of the variance in behaviour, if not more, than rivalling models of behaviour, including the theory of planned behaviour (27%; Armitage & Conner, 2001), theory of reasoned action (12.3%; Sheppard *et al.*, 1988), and the health belief model (0.5 to 4%; Harrison *et al.*, 1992). Further, both the CFA and EFA models showed superior fits for the multidimensional solution over the unidimensional solution. This is consistent with the broader COM-B literature, which proposes that each of the six subcomponents of COM individually and in interaction with one another are the key drivers of behaviour.

Fourth, the questionnaire was described by both health care professionals and people with low socio-economic status as being easy to understand, interesting, and well balanced. Reported dissatisfaction mainly concerned technical difficulties, formatting and layout, which were judged by the research team to be minor issues, and may have been caused due to viewing the questionnaire on a mobile device. However, findings must also be interpreted in the light of the differing levels of missing data across Time 1 and Time 2. At Time 1, there was less than 10% missing data across the six items. At Time 2, missing

data were higher (range = 18.1–32.5% across the six items). There were missing data on all six items for thirty participants. The item least likely to be completed was *social opportunity*; this finding occurred at both Time 1 and Time 2. Additionally, at Time 1 and Time 2, ceiling effects were observed for the same two items (*physical capability* and *psychological capability*). The precise reasons for the differences in missing data could be explained by the Time 1 questionnaire being incentivized (by the YouGov points system), but not the Time 2 questionnaire. In addition, the open-text comments suggested that some participants experienced technical difficulties completing the questionnaire at Time 2. Further testing beyond the present target behaviour is needed to assess whether levels of missing data differ across behaviours.

Strengths and limitations

This is the first study to describe the development and psychometric validation of a brief, generic measure of COM to assess perceived capabilities, opportunities, and motivations. Whilst we found evidence of acceptability, validity, and reliability, future testing is needed to examine further the psychometric properties of the questionnaire in the context of other behavioural domains and study populations to test the wider applicability of the questionnaire beyond the target behaviours in the present study. We recruited health care professionals and a general population sample with low socio-economic status in the development of the questionnaire. However, our final sample only involved health care professionals, and a caveat of these findings is that further testing is required among clinical and general population samples to determine its generalization. Specific areas for further research are to assess floor and ceiling effects over time (McHorney & Tarlov, 1995), and to examine specific reasons for missing data at Time 2, particularly in cases where participants are not incentivized to complete questionnaires.

With the limited number of studies deploying questionnaires specifically based on COM-B targeting health care professional practice, it was difficult to compare our item completion rates with any recognized benchmarks. Previous studies deploying COM-B questionnaires do not provide details of individual item completion rates (Ayton *et al.*, 2017; Balku *et al.*, 2017; Hankonen *et al.*, 2017; Stevely *et al.*, 2018; Taylor *et al.*, 2016; Webb *et al.*, 2016). Whilst our questionnaire completion rate (participants completing all 6 items) at Time 2 (61%) is comparable to the response rate achieved in the Ayton et al cross-sectional study (60%; Ayton *et al.*, 2017), we observed a number of 'don't know' responses at Time 1 (range 10.2%–22.5%). Given the analyses reported above, it is unlikely that our 'don't know' responses reflect question ambiguity, but may reflect ambivalence, satisficing, intimidation, or self-protection (e.g., Krosnick & Presser, 2010). It is notable, for example, that 'don't know' responses were highest for the physical and social opportunities items, and people *are* highly unlikely to know precisely what are their physical and social opportunities. In future use of our COM-B measure, we would recommend either: (1) exclude the 'don't know' option from the original questionnaire or (2) ensuring respondents who answer 'don't know' are asked follow-up exploratory questions (see Krosnick & Presser, 2010). In terms of future research, it would be valuable to identify precisely whether ambivalence, satisficing, intimidation, and/or self-protection are driving 'don't know' responses in our measure.

Additionally, given the high correlations observed between the pairs of each of the subdomains of COM (particularly in cases with variables correlating >.7 but <.85), further testing in additional samples is needed to help determine whether this pattern is consistently observed. Our *physical capability* item must also be subject to further

scrutiny, given the bias observed during our detailed analysis of the items, as well as the high number of 'don't know' responses to this item and our *social opportunity* item (20.7 and 22.5%, respectively). Making the questionnaire easily accessible with limited technical difficulties must also be a key aim of future iterations.

Conclusion

This study outlines the development and psychometric testing of the first brief, generic measure of COM to assess perceived capabilities, opportunities, and motivations. The COM-B is widely recognized in both public health guidelines (National Institute for Health & Care Excellence, 2014) and the literature (Michie *et al.*, 2014; Michie, van Stralen, et al., 2011) as providing the foundation for understanding behaviour change. Our questionnaire still has room for improvement, but with emphasis on its use in other settings and samples (with continued focus on acceptability, reliability, and validity data), we hope this provides a tool for policymakers and intervention developers to target known drivers to behaviour change, which can be adapted for use in different target behaviours.

Acknowledgements

This study was funded by a research grant obtained from Tesco Plc. Tesco had no role in the design of this study and did not have any role during its execution, analyses, interpretation, and storage of the data or decision to submit results.

Conflicts of interest

All authors declare no conflict of interest.

Author contribution

Chris Keyworth: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing; Tracy Epton: Investigation, Methodology, Writing – review & editing; Joanna Goldthorpe and Rachel Calam: Writing – review & editing; Christopher J. Armitage: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*, 471–499. https://doi.org/10.1348/014466601164939

Atkins, L., Kelly, M. P., Littleford, C., Leng, G., & Michie, S. (2017). Reversing the pipeline? Implementing public health evidence-based guidance in English local government. *Implementation Science*, *12*, 63. https://doi.org/10.1186/s13012-017-0589-5

Ayton, D. R., Barker, A. L., Morello, R. T., Brand, C. A., Talevski, J., Landgren, F. S., . . . Botti, M. (2017). Barriers and enablers to the implementation of the 6-PACK falls prevention program: A pre-implementation study in hospitals participating in a cluster randomised controlled trial. *PLoS ONE*, *12*, e0171932. https://doi.org/10.1371/journal.pone.0171932

Balku, E., Tóth, G., Nárai, E., Zsiros, E., Varsányi, P., & Vitrai, J. (2017). Methodology for identification of healthstyles for developing effective behavior change interventions. *Journal of Public Health*, *25*, 387–400. https://doi.org/10.1007/s10389-017-0799-y

Barker, F., Atkins, L., & de Lusignan, S. (2016). Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. *International Journal of Audiology*, *55*, S90–S98. https://doi.org/10.3109/14992027.2015.1120894.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160. https://doi.org/10.1177/096228029900800204

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155. https://doi.org/10.1037/0033-2909.112.1.155

Field, A. (2009). *Discovering statistics using SPSS*, 3rd edn. London: SAGE Publications Limited.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley Sons.

Giesinger, J. M., Kieffer, J. M., Fayers, P. M., Groenvold, M., Petersen, M. A., Scott, N. W., . . . EORTC Quality of Life Group. (2016). Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *Journal of Clinical Epidemiology*, *69*, 79–88. https://doi.org/10.1016/j.jclinepi.2015.08.007

Hankonen, N., Heino, M. T., Kujala, E., Hynynen, S. T., Absetz, P., Araujo-Soares, V., . . . Haukkala, A. (2017). What explains the socioeconomic status gap in activity? Educational differences in determinants of physical activity and screentime. *BMC Public Health*, *17*(1), 144. https://doi.org/10.1186/s12889-016-3880-5

Harrison, J. A., Mullen, P. D., & Green, L. W. (1992). A meta-analysis of studies of the Health Belief Model with adults. *Health Education Research*, *7*(1), 107–116. https://doi.org/10.1093/her/7.1.107

Howlett, N., Schulz, J., Trivedi, D., Troop, N., & Chater, A. (2017). A prospective study exploring the construct and predictive validity of the COM-B model for physical activity. *Journal of Health Psychology*, *24*(10), 1378–1391. https://doi.org/10.1177/1359105317739098

Keyworth, C., Epton, T., Goldthorpe, J., Calam, R., & Armitage, C. J. (2018). Are healthcare professionals delivering opportunistic behaviour change interventions? A multi-professional survey of engagement with public health policy. *Implementation Science*, *13*(1), 122. https://doi.org/10.1186/s13012-018-0814-x

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Krosnick, J. A., & Presser, S. (2010). Questionnaire design. In P. V. Marsden, & J. D. Wright (Eds.), *Handbook of survey research* (2nd edn., pp. 263–313). Bingley, UK: Emerald Group Publishing Limited.

McEachan, R., Taylor, N., Harrison, R., Lawton, R., Gardner, P., & Conner, M. (2016). Meta-analysis of the Reasoned Action Approach (RAA) to understanding health behaviors. *Annals of Behavioral Medicine*, *50*, 592–612. https://doi.org/10.1007/s12160-016-9798-4

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*, 293–307. https://doi.org/10.1007/BF01593882

Michie, S., Atkins, L., & West, R. (2014). *The behaviour change wheel: A guide to designing interventions*. London, UK: Silverback.

Michie, S., Van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*, 42. https://doi.org/10.1186/1748-5908-6-42

National Institute for Health and Care Excellence. (2014). *Behaviour change: Individual approaches (PH49)*. London, UK: NICE.

Plotnikoff, R. C., Costigan, S. A., Karunamuni, N., & Lubans, D. R. (2013). Social cognitive theories used to explain physical activity behavior in adolescents: A systematic review and meta-analysis. *Preventive Medicine*, *56*, 245–253. https://doi.org/10.1016/j.ypmed.2013.01.013

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research*, *15*, 325–343. https://doi.org/10.1086/209170

Simon, R., & West, R. (2015). Models of addiction and types of interventions: An integrative look. *The International Journal of Alcohol and Drug Research*, *4*(1), 8. https://doi.org/10.7895/ijadr.v4i1.198

Stevely, A. K., Buykx, P., Brown, J., Beard, E., Michie, S., Meier, P. S., & Holmes, J. (2018). Exposure to revised drinking guidelines and 'COM-B' determinants of behaviour change: Descriptive analysis of a monthly cross-sectional survey in England. *BMC Public Health*, *18*(1), 251. https://doi.org/10.1186/s12889-018-5129-y

Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics*, Vol. *5*. Boston, MA: Pearson.

Taylor, M. J., Arriscado, D., Vlaev, I., Taylor, D., Gately, P., & Darzi, A. (2016). Measuring perceived exercise capability and investigating its relationship with childhood obesity: A feasibility study. *International Journal of Obesity*, *40*(1), 34–38. https://doi.org/10.1038/ijo.2015.210

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Wamper, K. E., Sierevelt, I. N., Poolman, R. W., Bhandari, M., & Haverkamp, D. (2010). The Harris hip score: Do ceiling effects limit its usefulness in orthopedics? *Acta Orthopaedica*, *81*, 703–707. https://doi.org/10.3109/17453674.2010.537808

Webb, J., Hall, J., Hall, K., & Fabunmi-Alade, R. (2016). Increasing the frequency of physical activity very brief advice by nurses to cancer patients. A mixed methods feasibility study of a training intervention. *Public Health*, *139*, 121–133. https://doi.org/10.1016/j.puhe.2016.05.015

Young, M. D., Plotnikoff, R. C., Collins, C. E., Callister, R., & Morgan, P. J. (2014). Social cognitive theory and physical activity: A systematic review and meta-analysis. *Obesity Reviews*, *15*, 983–995. https://doi.org/10.1111/obr.12225

Received 2 October 2019; revised version received 23 March 2020

Supporting Information

The following supporting information may be found in the online edition of the article:

Supplementary Material File A. STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies.***

Supplementary Material File B. Phase 1: COM-B instrument design and development.

Supplementary Material File C. Questionnaire items (adapted; alternative text is presented in italics).

Supplementary Material File D. Figure 1. Bland-Altman plots to examine systematic differences between *Physical Opportunity* Time 1 and Time 2 measures. **Figure 2.** Bland-Altman plots to examine systematic differences between *Social Opportunity* Time 1 and Time 2 measures. **Figure 3.** Bland-Altman plots to examine systematic differences between *Reflective Motivation* Time 1 and Time 2 measures. **Figure 4.** Bland-Altman plots to examine systematic differences between *Automatic Motivation* Time 1 and Time 2 measures. **Figure 5.** Bland-Altman plots to examine systematic differences between *Physical Capability* Time 1 and Time 2 measures. **Figure 6.** Bland-Altman plots to examine systematic differences between *Psychological Capability* Time 1 and Time 2 measures. **Figure 7.** Heatmap demonstrating test-retest agreement scores for *physical opportunity*. **Figure 8.** Heatmap demonstrating test-retest agreement scores for *social opportunity*. **Figure 9.** Heatmap demonstrating test-retest agreement scores for *reflective motivation*. **Figure 10.** Heatmap demonstrating test-retest agreement scores for *automatic motivation*. **Figure 11.** Heatmap demonstrating test-retest agreement scores for *physical capability*. **Figure 12.** Heatmap demonstrating test-retest agreement scores for *psychological capability*.