

Running head: WHAT do SE items measure?

What do Self-Efficacy items measure?

Examining the Discriminant Content Validity of Self-Efficacy Items

Alison M.G. Burrell<sup>1</sup>, Julia L. Allan<sup>1</sup>, David M. Williams<sup>2</sup>, Marie Johnston<sup>1</sup>

1. Health Psychology, Institute of Applied Health Sciences, University of Aberdeen, U.K.

2. Behavioral and Social Sciences, School of Public Health, Brown University, U.S.A.

Correspondence concerning this paper should be sent to:

Dr Julia Allan, Health Psychology, Institute of Applied Health Sciences, 2<sup>nd</sup> Floor Health Sciences Building, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD United Kingdom. Tel: 01224 438103. Email: [j.allan@abdn.ac.uk](mailto:j.allan@abdn.ac.uk)

## **Abstract**

**Objectives.** Self-efficacy – an individual’s judgement of their ability to successfully perform a behaviour- is commonly used to explain and predict behaviour. It is measured through self-report questionnaires. These scales require good content validity, i.e. must measure the full scope and content of the construct without contamination from similar constructs. The present study uses a systematic, transparent quantitative method (Discriminant Content Validation) to assess the content validity of a variety of self-efficacy items and qualitatively explores participant interpretations of these items.

**Design.** A quantitative Discriminant Content Validation (DCV) and qualitative think aloud study of self-efficacy item interpretation.

**Methods.** Participants (n=21) were presented with items designed to measure self-efficacy and related constructs following standard DCV methodology. Items were rated against construct definitions to determine if they measured a particular construct (Yes/No). Judges’ confidence in each assessment was also assessed (%) and used to establish quantitative estimates of content validity for each item. A qualitative think aloud study explored the judgements made in a subset of participants.

**Results.** 8/8 self-efficacy items were found to measure self-efficacy, however, 2/8 of these also measured motivation. 6/8 items displayed discriminant content validity and thus can be considered ‘pure’ measures of self-efficacy. The think aloud study indicated that item wording is a likely cause of item misinterpretation.

**Conclusions.** Self-efficacy items vary in terms of their content validity with only some of the items assessed providing ‘pure’ measures of the self-efficacy construct. Item wording should be considered during study design to avoid misinterpretation.

*Keywords:* self-efficacy, discriminant content validity

## Introduction

Self-efficacy is defined by Bandura as an individual's judgement of their ability to successfully perform a behaviour (Bandura, 1986). This can be a belief in a physical or mental skill or a more complex belief about perceived demands or impediments (Bandura, 1997). As a primary construct of social cognitive theory, self-efficacy is one of the constructs most frequently used to predict, explain and change behaviours (Glanz & Bishop, 2010) such as physical activity, smoking, healthy eating and alcohol abstinence (Williams & Rhodes, 2016). It is also an influential construct in other health behaviour theories and frameworks such as the Health Action Process Approach (Schwarzer, 1992; 2008) and is conceptually similar to (although distinguished from) perceived behavioural control in the Theory of Planned Behaviour (Ajzen, 2002) or the concept of capability in the COM-B system (Michie, van Stralen & West, 2011). Indeed, at present there are over 46,600 articles citing the term 'self-efficacy' in the PubMed data base (US National Library of Medicine, 12 May, 2017). As such an important construct, it is vital that all aspects of the validity of self-efficacy measures are reliably established.

**Measuring self-efficacy.** Bandura (2006) recommends that self-efficacy items should be phrased in terms of *can* or *could do* rather than *will do* statements, ensuring that a person's *intention* to carry out a behaviour and their *perceived capability* to carry out a behaviour remain separate both conceptually and empirically (p.309). Any items using the scale markers 'can', 'could', 'ability to' or 'confidence to' are considered by other authors to be consistent with Bandura's definition of self-efficacy and guide to measuring it (William & Rhodes, 2016). A previous study using DCV methods demonstrated that commonly used self-efficacy items do indeed measure self-efficacy as defined and are not contaminated by other perceived control constructs (Johnston et al., 2014). However, while self-efficacy items have been found to be distinguishable from similar constructs from other theories, it is

Running head: WHAT do SE items measure?

equally important to ensure that they are uncontaminated by different constructs within the same theory. Otherwise, when the theory is used to investigate behaviour, variance attributable to self-efficacy may be attributed to a different construct or vice versa.

Despite its unwavering popularity in health behaviour theories and interventions, questions have been raised about the measurement of Bandura's original concept and the potential confounding between self-efficacy, motivation and outcome expectancies (Corcoran, 1991; Kirsch, 1982, 1995; Williams, 2010) with particular focus on the scales used to measure self-efficacy (Rhodes & Blanchard, 2007; Rhodes, Williams & Mistry, 2016; Williams & Rhodes, 2016).

### **Self-efficacy-as-motivation.**

“Could you eat a live worm? Could you laugh out loud during the middle of a funeral? Could you kill a baby kitten?...Clearly when you say you cannot do these things, you mean something different than when you say you cannot solve a difficult calculus problem, lift a 300-pound weight, or successfully execute the job requirements of an astronaut” (Kirsch, 1995, p. 338-339).

Kirsch alludes to the idea that participants may respond to such questions based on *motivation* to avoid revulsion, embarrassment, shame or feelings of guilt (as cited in Williams & Rhodes, 2016, p. 117) rather than *perceived ability*. While Bandura's writing about the conceptual distinction between self-efficacy and motivation is unclear, it is likely that the way in which measurement items are phrased will affect whether or not responses capture one or both constructs. Participants responding to the item 'could you eat a live worm?' may or may not unconsciously add a conditional 'if' (e.g. 'if I had to'). In three studies conducted by Rhodes and Blanchard (2007) and Rhodes and Courneya, (as cited in Williams & Rhodes,

Running head: WHAT do SE items measure?

2016), an attempt was made to remove ambiguity from items in self-efficacy scales by adding the phrase '*if you really wanted to*' to each item. In all three studies, 'if you really wanted to' items were judged as measuring perceived ability more accurately, suggesting the original items are open to alternative interpretation and may partially reflect other processes such as motivation or social norms (p. 120). For example, a participant asked to complete an item measuring their self-efficacy to exercise regularly when it is raining may respond with a rating of 50 (moderately can do) when presented with a standard item but increase their rating to 90 (certain can do) when the item is appended with '*if I really wanted to*', indicating that the standard item is measuring something more than just perceived ability – perhaps their desire not to get wet. This argument led Williams and Rhodes (2016) to propose that self-efficacy scale items capture both perceived ability and motivation. That is, that items may not measure only the construct they are designed to measure.

The present study uses the Discriminant Content Validity method developed by Johnston et al. (2014) to formally evaluate the precision with which commonly used self-efficacy items reflect the self-efficacy construct. DCV methods have been successfully used to assess the content of measures assessing: illness representations (Johnston et al., 2014), theoretical domains (Huijg, Gebhardt, Crone, Dusseldorp, & Plesseau, 2014), health outcomes (Pollard, Johnston & Dieppe, 2006), pain (Dixon, Pollard & Johnston, 2007), automaticity (Gardner, Abraham, Lally, & de Bruijn, 2012) and work stress constructs (Bell, Johnston, Allan, Pollard & Johnston, 2017). In each case, items which are pure measures of each of the theoretical constructs have been identified in addition to items contaminated with content from other theoretical constructs and items which fail to measure the intended construct.

In the present study, standard self-efficacy items taken from published scales were presented to members of the general public to determine whether they are perceived by

Running head: WHAT do SE items measure?

participants to be relevant to self-efficacy as defined by Bandura. More specifically, the study aims to investigate whether there is a difference between people's interpretation of items that vary in their wording ('certain' vs 'confident', 'can', 'could', 'could if I wanted to', 'am capable of').

## Methods

**Design.** The present study follows the standard 6-step discriminant content validation process, as outlined by Johnston et al. (2014) followed by a think aloud study. Following peer review of the study's protocol, ethical approval was applied for and granted by the relevant University Ethics Committee.

**Participants.** Participant characteristics are summarised below in Table 1. Participants (n=21) were recruited using social media and word-of-mouth, from outside the field of psychology to avoid participants with detailed prior knowledge of the self-efficacy construct. All participants were over 18 and were native English speakers. Before the study was conducted participants were given an information sheet regarding the nature of the study which was described as investigating how questionnaire items are interpreted by the general public. Participants then provided informed consent to participate, with the option to opt-out of the think-aloud study.

**Step 1. Identify clear definitions of constructs.** Based on past attempts to disentangle the construct of self-efficacy, four constructs were selected for the DCV: *self-efficacy* itself, and three other constructs which may potentially overlap with self-efficacy: *motivation* (identified by Williams & Rhodes, 2016 as a construct which may be inadvertently captured by self-efficacy items), *outcome expectancies* (an important construct in social cognitive theory and identified by Williams (2010) as a construct that judgements

Running head: WHAT do SE items measure?

about self-efficacy may be based upon) and *opportunity* (a construct potentially related to self-efficacy as perceptions of actual control may influence perceptions of perceived control as in the Theory of Planned Behaviour). Definitions (Table 2) for the constructs to be included in the DCV were extracted from the literature by the research team.

Original definitions by Bandura (1986) were used for *self-efficacy* and *outcome expectancies*. The construct *motivation* proved to be more difficult to define, with a multitude of differing definitions in published literature. Many of these were felt to be overly complex for non-academic participants to understand or not inclusive of the full scope of the construct. Following a research team discussion, Baumeister & Vohs's (2007) definition of motivation was used. Michie, van Stralen and West's (2011) definition of *opportunity* was selected from their COM-B model.

**Step 2. Item generation.** An extensive literature review was carried out to find scales used to measure self-efficacy (Figure 1.). In order to provide direction and structure, this literature review was restricted to studies using self-efficacy to predict one common health behaviour - physical activity. Three recent systematic reviews were identified and used to locate original empirical studies which measured self-efficacy to participate in physical activity (Ashford, Edmunds & French, 2010; Orlander et al., 2013; Prince et al., 2016). The most common phrasing observed across the 131 unique items identified used 'certain' and 'confident' in combination with the verbs 'can' and 'could' and the phrase 'capable of'. These variations were selected for inclusion in the DCV along with items extended with 'if I wanted to' to test the 'self-efficacy-as-motivation' hypothesis proposed by Williams & Rhodes (2016). Eight variations of item wording and response scales were selected for DCV; 1) certain I am capable of.....; 2) certain I could.....; 3) certain I can.....; 4) confident I am capable of.....; 5) confident I could.....; 6) confident I can.....; 7) certain I could.....if I wanted to; and 8) confident I can.....if I wanted to. In order to encourage participants to

Running head: WHAT do SE items measure?

focus on the constructs and item wording rather than a particular behaviour, and for generalisability, all reference to physical activity was replaced with 'BEHAVIOUR X'.

The literature was also searched for scales measuring *motivation, outcome expectancies* and *opportunity* based on the definitions chosen. N=63 items were identified from this search and presented to two independent raters who selected the items that they felt best measured each construct definition. Finally, two 'other' items were chosen. These items were selected as items that definitely did NOT measure any of the constructs to be assessed in order to assess participants' general ability to assign items to definitions, as recommended by Johnston et al. (2014). These 'other' items measured personality and cognitive performance. Full versions of all items included in the DCV are listed in Supplementary File S2.

**Step 3. Identify appropriate judges** The DCV 'judges' were recruited from the general population as they represent the populations included in self-efficacy studies so can be used to investigate whether lay interpretations of items and response scales were as intended by researchers developing self-efficacy scales. Johnston et al. (2014) recommends at least 15 judges in order to establish statistically significant content validity. Participants were required to have a proficient level of English and to not have a university level psychology background.

**Step 4. Establish a scale.** The construct definitions from Step 1 were then combined with the 16 selected questionnaire items from Step 2 to form a DCV questionnaire (Figure 2). In order to prevent bias, the labels were removed from the construct definitions. Self-efficacy was labelled 'A', motivation 'B', outcome expectancies 'C', and opportunity 'D'. 'Other' was included as category 'E' to provide participants with the option to indicate that an item presented measured something other than constructs A-D. Participants were asked to decide



Running head: WHAT do SE items measure?

whether each item measured each of the constructs (yes/no) and to rate their confidence in each decision (0-100%).

Participants' confidence judgements were then divided by 100 (and multiplied by -1 if they indicated that the item was *not* measuring a construct), resulting in a scale from -1 (confidently does NOT measure a construct) to +1 (confidently DOES measure a construct). This data was input into SPSS v22 for analysis.

**Step 5. Test the content validity.** To test the content validity of each questionnaire item in relation to each theoretical construct, Wilcoxon one-sample tests were carried out for each item for each construct. Significant ( $p < .01$ ) results indicated that participants were confident that the questionnaire item was (positive score) or was not (negative score) measuring the construct in question. A Bonferroni correction was used to correct for the analysis of multiple items.

**Step 6. Evaluate the DCV.** Each item selected for the DCV came from a particular scale and was considered by the research team to measure only one construct in accordance with the definitions selected. Therefore, items should only have content validity for one construct. Showing validity for more than one item indicates that contamination is present and that the item in question does not have discriminant content validity.

**Think-aloud Methods** As suggested by Johnston et al. (2014) a follow-up think aloud study was then conducted in a randomly selected group of participants from the main DCV study ( $n=4$ ). Following a brief pilot test, participants were presented with the items purporting to measure self-efficacy and asked to verbalise their thoughts on the wording of the items and their interpretation of the items. This data was qualitatively summarised using basic thematic analysis (Braun & Clarke, 2006).

Running head: WHAT do SE items measure?

## Results

Results from single sample Wilcoxon signed ranks tests are outlined in Table 3.

**Do the questionnaire items included in the DCV measure the constructs they purport to measure?** 8/8 self-efficacy items were judged by participants to measure the self-efficacy construct as outlined by Bandura (1986). 2/2 *motivation* items, 2/2 *outcome expectancies* items and 1/2 *opportunity* items were judged to be measuring their intended constructs. In contrast, neither of the *other* items were correctly identified as measuring something other than the constructs provided.

**Do the questionnaire items measure a construct(s) that they are not intended to measure?** 2/8 self-efficacy items (*I am certain that I could...* and *I am certain that I could if I wanted to*) were perceived to be a measure of motivation as well as a measure of self-efficacy. One *other* item (*I am the kind of person who does...*) was perceived by participants to be a measure of motivation.

**Do the items which were correctly allocated to the construct they purport to measure have discriminant content validity (DCV)?** Items were considered by researchers to have discriminant content validity if they were correctly allocated to the construct they are intended to measure, and only that construct. 6/8 self-efficacy items were shown to have DCV, while 2/2 *motivation* items and 2/2 *outcome expectancies* items displayed DCV. 1/2 *opportunity* items displayed DCV.

**Estimate of effect sizes.** In order to estimate the size of the effects observed in the DCV, rank correlations for each self-efficacy item were calculated (a form of Spearman's  $r$  calculated from the Wilcoxon test statistic,  $W$ , divided by the total rank sum) and are illustrated in Table 4. Following Bell et al. (2017) we used a threshold of .8 correlation (equivalent to 64% shared/not shared content) to indicate that items significantly shared

Running head: WHAT do SE items measure?

content with the target construct. Significant rank correlations of  $>.8$  level were observed for all self-efficacy items, except 'certain I could'. Indeed, 'certain I could' showed a stronger rank correlation with motivation (.94) than self-efficacy (.7)

### **Think aloud results**

Data from the think aloud study was transcribed and coded for emergent themes, in order to gain a deeper understanding of the results observed in the DCV study. Five main findings were apparent;

**Finding 1:** Participants commented that the wording of certain items alluded or partly alluded to motivations to carry out a behaviour, rather than a judgement of ability. They particularly focused on the phrase 'could get yourself to...' as a measure of drive or inclination, included in Bandura's (2006) guide.

*"Some of these I felt did not really measure whether a person could successfully do something or not, but whether I could do it or had an inclination to do it" (P123).*

*".... 'could get yourself' to do something is almost implying that you could possibly see yourself doing this, you know, rather distasteful act or something. [laughs] It's a bit strange this whole could get yourself, it's almost implying, could you overcome your inner revulsion at this, em...so I didn't really find that this in any way getting at the ability thing" (P123).*

*"Get yourself. So, get yourself...like that's more to do with the motivation. You know, behind it. Em, like more to do with the B [motivation]. The drive or the inclination rather than actually the ability to be able to do it" (P115).*

**Finding 2:** Participants differed over their interpretation of ability. While some considered perceived ability as an act that they would be possibly able to try, others envisaged it as a behaviour for which they already have a particular skill or level of competency.

*“... capable here seems to suggest, do you have a minimum ability to get by at this, in my understanding of it, it doesn’t seem to say ‘I am confident that I can, you know, successfully do this task’ (P123).*

*“Again, the word ‘certain’ is pretty definite and sure and then capable that I have the skills, the ability or the training” (P115).*

**Finding 3:** Participants interpreted the word ‘could’ in different ways, as a possibility in a hypothetical situation and as the past of the verb ‘can’.

*“I could do it. In that, it wouldn’t be something beyond the realm of possibility. Rather than is it within your capabilities” (P123)*

*“‘Could not do at all’ and ‘could do’ at 100. Em, just it seems like it’s asking you what you could do in the past and not at present. Not a present study of ability” (P111).*

**Finding 4:** Participants reported greater understanding that items measured ability when the word ‘confident’ was used (rather than ‘certain’).

Running head: WHAT do SE items measure?

*“I am confident that I could perform yoga. I’m very confident. Again, that seems like my confidence...my degree of confidence in my own ability” (P123)*

*“I am confident that I could perform the behaviour. So again, that’s more to do with the ability, that I have whatever skills or whatever you need to be able to actually do it” (P115)*

*“I am confident...It’s more to do with my ability” (P124)*

**Finding 5:** The phrase ‘If I wanted to’ was interpreted by participants as a way of including motivation in the item.

*“And I’m confident that I could perform behaviour X...if I wanted to. Oh ok so this is the one with ‘if I wanted to’ here. It doesn’t really change it at all for me. My mind is drawn to this ‘could get yourself’ thing. I’m confident that I could perform, say yoga, if I wanted to. Again that’s actually testing ‘can you overcome preconceived ideas that you have about this and make yourself do this rather than can you be good at it? So yeah, that’s it” (P123).*

*“The ‘if I wanted to’ kind of makes me a bit towards B [motivation]” (P124)*

*“It’s got the... ‘if I wanted to’ so, if you’re actually motivated to do it or not” (P115)*

## **Discussion and Conclusion**

The discriminant content validity of items commonly used to measure self-efficacy was assessed to ascertain how such items are interpreted by intended respondents.

The results show that all eight of the items previously used in research to measure self-efficacy were perceived by participants as a measure of the construct, with six items considered to be ‘pure’ measures, that is, demonstrating discriminant content validity. However, two of these items, were found to also measure motivation. These items both used the wording ‘...*certain I could*...’ with one also including the phrase ‘...*if I wanted to*...’. These findings are congruent with the suggestion made by Williams and Rhodes (2016) that some self-efficacy items are indeed not perceived by participants in the way intended by researchers. Use of these contaminated items might give misleading results, especially in a study where both self-efficacy and motivation were relevant constructs.

These results have two main implications. First, in designing a study which includes measures of self-efficacy, care should be taken to include pure measures by either using wording shown here to have DCV or by examining the DCV of items selected or developed. It is recommended based on the present results, that items with the structure ‘certain I could X’ be avoided when assessing self-efficacy as these are not pure measures of the construct. In addition, the results reported here suggest wording which might be easier for respondents to understand. Previous studies have identified pure items for other theoretical constructs including disability and work stress and recommended pure items. For example, in studies of disability, the relationship between impairment and disability is frequently investigated but often contaminated measures of either construct are used and can produce misleading results (Johnston & Pollard, 2001). Second, published studies use a variety of measures to assess self-efficacy and our results suggest that these need to be interpreted with caution. Some of the items used may be contaminated by other constructs such as motivation and the think aloud study suggests which wordings may be most problematic.

Results support Kirsch’s (1995) suggestion that the use of the word ‘could’ can lead to misinterpretation and is situation dependent. He argues that using ‘could’ in self-efficacy

Running head: WHAT do SE items measure?

items can lead to a variety of interpretations related to ability, avoidance of embarrassment, revulsion or feelings of guilt. This may provide a reason why two of the items assessed were rated as measuring perceived ability and motivation equally. Participants in the think aloud study commented on the use of the word ‘could’ and in particular, on the phrase provided by Bandura (1996), ‘...could get yourself to...’

In order to resolve the issue of contaminated self-efficacy items which also measure motivation, Williams and Rhodes (2016) suggested adding the phrase ‘*if I wanted to*’ to self-efficacy items to remove such ambiguity. One of the two items which included ‘*if I wanted to*’ however, was perceived in the present study as *measuring* motivation, rather than *removing* motivation. It is important to note that the ‘*if I wanted to*’ item that was perceived as measuring motivation also included the ‘*could get yourself to*’ phrase. Thus, the confounding of this item with motivation could be a function of the latter phrase rather than the ‘*if I wanted to*’ phrase. Nonetheless, this finding was surprising given that the ‘*if I wanted to*’ phrase should, if anything, further parse motivation from perceived capability rather than confound the two concepts. Indeed, any question with a suffix of the form ‘*if X*’ should serve to logically control for the effects of X so as to remove variability in X from consideration in response to the question. However, the qualitative findings showed that this linguistic logic clearly was not interpreted in this way. The item in question ‘certain...you could get yourself to . . . if you wanted to’ is therefore contradicting the studies previously carried out by Rhodes and Blanchard (2007) and Rhodes and Courneya, (as cited in Williams & Rhodes, 2016). This highlights the fact that respondents’ interpretation of questionnaire items may be very different—from their literal meanings and from researchers’ intentions – and therefore items need to be selected or investigated with care.

In contrast, six of the items were found to measure self-efficacy without contamination from other constructs and were therefore deemed to be ‘pure’ self-efficacy

Running head: WHAT do SE items measure?

items, illustrating DCV. These items are therefore usable in studies including both self-efficacy and some of the other constructs investigated here. Nevertheless, it is possible that they might still be contaminated by constructs not investigated here, for example by other constructs related to beliefs in control. A previous study comparing self-efficacy items to perceived behavioural control items (Johnston et al., 2014) tested only items in the format “I can....X” so it remains to be established whether items with other phrasing are equally discriminable from related control constructs. The most highly rated items used the words ‘capable of’, ‘certain’ and ‘confident’. Indeed, think aloud participants commented on the use of ‘confident’, as easier to interpret. Participants reported that they found some of the items, particularly those in the format outlined by Bandura (2006) to be complex and difficult to interpret.

**Study limitations** Although an extensive review of literature was carried out before the DCV was conducted, the list of constructs which are possible contaminants is not exhaustive. Other related constructs such as self-esteem could be examined in the future. The construct definitions used in the current study are based on researcher consensus. As observed, researchers’ perceptions may not always reflect what a participant perceives. The *motivation* construct proved to be particularly challenging, as it is a word commonly used in everyday speech with a multitude of definitions.

As highlighted by Kirsch (1995), the perception of certain words and items can be dependent on the situation, behaviour, and population. While specific behaviours in the present study were replaced with ‘behaviour X’ to establish a clearer focus on the item wording used, in practice, the target behaviours in questionnaires will vary in difficulty and skill (e.g. walk 100m versus run a marathon). Most importantly, Kirsch (1995) argued that the words “can” or “could” will be interpreted differently if they are in reference to skilled behaviours, such as shooting basketball hoops or solving calculus problems, than if they are



Running head: WHAT do SE items measure?

in reference to choice behaviours, such as going for a walk or going to the cinema (for supporting empirical findings see Corcoran & Rutledge, 1989; Kirsch, 1982). For example, if a person's friend asks her if she "can" go to the cinema this evening, the person will interpret the question in terms of motivation, considering incentives and disincentives (e.g., wanting to spend time with her friend, but too much work to do, and a desire to get to bed early, etc.) rather than literal ability to go to the cinema. In the present study, participants were asked to determine whether questionnaire items assessing whether one thinks that they "can" perform behaviour X reflects their perceived "ability" to do so. Taken *literally*, the answer is yes. But if "behaviour X" was replaced with a choice behaviour such as going to the cinema, going for a walk, or exercising on a regular basis, then the response may have been different. The effects of this variation in behavioural context could not be assessed in the present study. However, qualitative finding #2 suggests that there is considerable variability in how respondents interpret the meaning of the words "ability" and "can" which suggests that this issue is worth investigating in future. Specifically, participants varied in whether they interpreted 'ability' as referring to a minimal basic ability (i.e. could attempt X) or as an indication that a certain level of skill was already present (i.e. certainly have the ability to do X). Future studies should investigate whether this interpretation varies systematically along with the type / difficulty of the behaviour in question

Finally, the majority of the participants recruited in the present study were educated to a university level, and may therefore not have been representative of the general population in their interpretation / reading of items.

**Conclusion** To sum up, while all self-efficacy items assessed were found to measure the construct, two were also found to capture motivation and could therefore not be considered 'pure', uncontaminated measures of self-efficacy. The results suggest that researchers should carefully consider the wording of the self-efficacy items they use and should take steps to

Running head: WHAT do SE items measure?

clarify the intended meaning of items or select the wordings found here to give pure measures with minimal respondent misinterpretation.

## References

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32*(4), 665-683. doi: 10.1111/j.1559-1816.2002.tb00236.x

Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology, 15*(2), 265-288. doi: 10.1348/135910709X461752

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T.C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information Age.

Baumeister, R. F., & Vohs, K. D. (2007). Self-Regulation, ego depletion, and motivation. *Social and Personality Psychology Compass, 1*(1), 115-128. doi: 10.1111/j.1751-9004.2007.00001.x

Running head: WHAT do SE items measure?

Bell, C., Johnston, D., Allan, J., Pollard, B., & Johnston, M. (2017). What do Demand-

Control and Effort-Reward work stress questionnaires really measure? A discriminant content validity study of relevance and representativeness of measures. *British Journal of Health Psychology*, 22(2), 295-329. doi: 10.1111/bjhp.12232

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research*

*in Psychology*, 3(2), 77-101. doi: 10.1191/1478088706qp063oa

Corcoran, K. J., & Rutledge, M. W. (1989). Efficacy expectation changes as a function of

hypothetical incentives in smokers. *Psychology of Addictive Behaviors*, 3, 22-29. doi: 10.1037/h0080563Issn

Corcoran, K. J. (1991). Efficacy, "skills," reinforcement and choice behaviour. *American*

*Psychologist*, 46(2), 155-157. doi: 10.1037/0003-066X.46.2.155

Dixon, D., Pollard, B., & Johnston, M. (2007). What does the chronic pain grade

questionnaire measure?. *Pain*, 130(3), 249-253. doi: 10.1016/j.pain.2006.12.004

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G. J. (2012). Towards parsimony in habit

measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 102. doi: 10.1186/1479-5868-9-102

Glanz, K., & Bishop, D. B. (2010). The role of behavioral science theory in development and

implementation of public health interventions. *Annual Review of Public Health*, 31, 399-418. doi: 10.1146/annurev.publhealth.012809.103604

Huijg, J. M., Gebhardt, W. A., Crone, M. R., Dusseldorp, E., & Pesseau, J. (2014).

Discriminant content validity of a theoretical domains framework questionnaire for use in implementation research. *Implementation Science*, 9(1), 11. doi:10.1186/1748-5908-9-11

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014).

Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *British Journal of Health Psychology*, 19(2), 240-257. doi: 10.1111/bjhp.12095

Johnston, M., & Pollard, B. (2001). Consequences of disease: testing the WHO International

Classification of Impairments, Disabilities and Handicaps (ICIDH) model. *Social Science & Medicine*, 53(10), 1261-1273. doi: 10.1016/S0277-9536(00)00384-1

Kirsch, I. (1982). Efficacy expectations or response predictions: the meaning of efficacy

ratings as a function of task characteristics. *Journal of Personality and Social Psychology*, 42(1), 132-136.

Kirsch, I. (1995). Self-efficacy and outcome expectancy: A concluding commentary. In J.E.

Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 341-345). New York, NY: Plenum Press.

Running head: WHAT do SE items measure?

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(1), 42. doi: 10.1186/1748-5908-6-42

Olander, E. K., Fletcher, H., Williams, S., Atkinson, L., Turner, A., & French, D. P. (2013).

What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis.

*International Journal of Behavioral Nutrition and Physical Activity*, 10(1), 29.

doi:10.1186/1479-5868-10-29

Pollard, B., Johnston, M., & Dieppe, P. (2006). What do osteoarthritis health outcome instruments measure? Impairment, activity limitation, or participation

restriction?. *The Journal of Rheumatology*, 33(4), 757-763.

Prince, S. A., Reed, J. L., Martinello, N., Adamo, K. B., Fodor, J. G., Hiremath, S.,

Kristjansson, E.A., Mullen, K.A., Nerenberg, K.A., Tulloch, H.E. & Reid, R. D.

(2016). Why are adult women physically active? A systematic review of prospective cohort studies to identify intrapersonal, social environmental and physical environmental determinants. *Obesity Reviews*, 17(10), 919-944. doi:

10.1111/obr.12432

Rhodes, R.E. & Blanchard, C.M. (2007). What do confidence items measure in the physical

Running head: WHAT do SE items measure?

activity domain? *Journal of Applied Social Psychology*, 37(4), 759-774. doi:  
10.1111/j.1559-1816.2007.00184.x

Rhodes, R. E., Williams, D. M., & Mistry, C. D. (2016). Using short vignettes to disentangle perceived capability from motivation: a test using walking and resistance training behaviors. *Psychology, Health & Medicine*, 21(5), 639-651. doi:  
10.1080/13548506.2015.1074710

Schwarzer R. (1992) Self-efficacy in the adoption and maintenance of health behaviors: theoretical approaches and a new model. In R. Schwarzer (Ed.), *Self-Efficacy: Thought Control of Action* (217-242). Washington, DC: Hemisphere.

Schwarzer, R. (2008). Modeling health behaviour change: how to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology* 57(1), 1-29. doi:  
10.1111/j.1464-0597.2007.00325.x

US National Library of Medicine. (2017, 12 May). *PubMed.gov*. Retrieved from  
<https://www.ncbi.nlm.nih.gov/pubmed/?term=self-efficacy>.

Williams, D. M. (2010). Outcome expectancy and self-efficacy: Theoretical implications of an unresolved contradiction. *Personality and Social Psychology Review*, 14(4), 417-425. doi: 10.1177/1088868310368802

Williams, D. M., & Rhodes, R. E. (2016). The confounded self-efficacy construct: conceptual analysis and recommendations for future research. *Health Psychology Review*, 10(2), 113-128. doi: 10.1080/17437199.2014.941998

**Tables**

Table 1. *Descriptive statistics for participants*

---

	n (%)
Age	
18-25	1(4.8)

---

Running head: WHAT do SE items measure?

26-35	14(66.7)
36-45	4(19)
46-55	1(4.8)
56-65	1(4.8)
Gender	
Male	9(43)
Female	12(57)
Highest level of education	
Secondary school	2(9.5)
Bachelor's degree	10(47.6)
Postgrad qualification	9(42.9)

Table 2. *Definitions of constructs and sources included in DCV task*

---

<u>Construct</u>	<u>Definition</u>	<u>Reference</u>
<b>Self-efficacy</b>	An individual's judgement of their ability to successfully perform a behaviour	Bandura (1986)
<b>Outcome</b>	The assessment a person makes	Bandura (1986)



Running head: WHAT do SE items measure?

**expectancies** that a particular behaviour will result  
in a particular outcome

**Motivation** Any sort of general drive or inclination to do something  
Baumeister & Vohs (2007)

**Opportunity** The factors that lie outside the individual (social and environmental) that make the behaviour possible or prompt it  
Michie, van Stralen & West (2011)

Table 3. *Wilcoxon Signed Ranks Test scores (Z and 2-tailed p value) ranked in order of measurement of intended construct. + value means item measures construct, - value means item does not measure construct. Items highlighted in bold reached statistical significance. Items in italics have been judged as measuring an incorrect construct.*

Results from DCV of *self-efficacy* items.

Item wording	Self-efficacy	Motivation	Outcome expectancies	Opportunity	Other
certain...capable of	<b>+4.413</b> <b>&lt;.001</b>	+ .891 .373	-.562 .574	-1.762 .078	<b>-3.413</b> <b>.001</b>
confident...capable of	<b>+4.289</b> <b>&lt;.001</b>	+ .466 .641	-2.001 .045	-1.742 .081	<b>-3.191</b> <b>.001</b>
confident...can	<b>+4.240</b> <b>&lt;.001</b>	+1.273 .203	<b>-2.588</b> <b>.010</b>	-1.087 .277	<b>-2.979</b> <b>.003</b>
confident...could	<b>+4.203</b> <b>&lt;.001</b>	+ .694 .487	-1.577 .115	-1.125 .261	<b>-3.380</b> <b>.001</b>

Running head: WHAT do SE items measure?

certain...can	<b>+4.162</b> <b>&lt;.001</b>	+1.070 .285	<b>-2.849</b> <b>.004</b>	-1.442 .149	<b>-3.751</b> <b>&lt;.001</b>
confident...can if I wanted to	<b>+3.749</b> <b>&lt;.001</b>	+1.743 .081	-1.801 .072	-1.276 .202	<b>-2.606</b> <b>.009</b>
certain...could if I wanted to	<b>+3.147</b> <b>.002</b>	<b>+3.165</b> <b>.002</b>	-1.700 .089	-.930 .352	-1.900 .057
certain...could	<b>+3.030</b> <b>.002</b>	<b>+3.705</b> <b>&lt;.001</b>	-1.949 .051	+.918 .359	<b>-2.958</b> <b>.003</b>

Table 3 (continued)

Results from DCV of *motivation* items

	Self- efficacy	Motivation	Outcome expectancies	Opportunity	Other
I...because I simply enjoy it	-1.727 .084	<b>+3.948</b> <b>&lt;.001</b>	-1.179 .239	<b>-3.454</b> <b>.001</b>	-2.462 .014
Most of the time, I feel I really want to do...	-2.040 .041	<b>+3.521</b> <b>&lt;.001</b>	<b>-3.113</b> <b>.002</b>	-1.853 .064	-1.615 .106

Results from DCV of *outcome expectancies* items

	Self- efficacy	Motivation	Outcome expectancies	Opportunity	Other
...would help me control my weight	-.991 .322	+.113 .910	<b>+4.177</b> <b>&lt;.001</b>	+.612 .540	<b>-2.867</b> <b>.004</b>
...would help me reduce tension and manage stress	-.466 .641	+1.595 .111	<b>+4.127</b> <b>&lt;.001</b>	-.566 .572	+2.095 .036

Results from DCV of *opportunity* items

	Self- efficacy	Motivation	Outcome expectancies	Opportunity	Other
...many places where I can...at home & work	-.131 .896	+.242 .809	<b>-2.678</b> <b>.007</b>	<b>+3.493</b> <b>&lt;.001</b>	-2.435 .015
Most people I know do...often	<b>-3.040</b> <b>.002</b>	+.061 .952	<b>-3.237</b> <b>.001</b>	+.626 .532	+1.445 .149

Table 3 (*continued*)

Results from DCV of *other* items

	Self- efficacy	Motivation	Outcome expectancies	Opportunity	Other
I am the kind of person who does...	-.868 .385	<b>+3.265</b> <b>.001</b>	-1.720 .085	-2.041 .041	+.075 .940
I need to concentrate to do...	+2.285 .022	+1.396 .163	+1.018 .309	-1.818 .069	-1.657 .098

	Self- efficacy	Motivation	Outcome expectancies	Opportunity	Other
certain...capable of	<b>+1*</b>	+0.22	-0.14	-0.46	<b>-.9</b>
confident...capable of	<b>+1*</b>	+0.12	-0.52	-0.44	<b>-.83</b>
confident...can	<b>+1*</b>	+0.32	-0.67	-0.28	-0.76
confident...could	<b>+1*</b>	+0.18	-0.41	-0.29	<b>-.86</b>
certain...can	<b>+1*</b>	+0.28	<b>-.74</b>	-0.39	<b>-.97</b>
confident...can if I wanted to	<b>+.92*</b>	+0.42	-0.46	-0.33	-0.68
certain...could if I wanted to	<b>+.8</b>	<b>+.77</b>	-0.43	-0.24	-0.49
certain...could	+0.7	<b>+.94</b>	-0.5	+0.23	<b>-.81</b>

Table 4. *Effect sizes (rank correlations) for self-efficacy items and each theoretical construct*

Items which showed significant results are presented in **bold**. +ve value indicates item is measuring construct. -ve value indicates item is not measuring construct. Results marked with an \* were found to show discriminant content validity.

## **Supplementary File S1**

### **Figures**

Figure 1. *Process of generating self-efficacy items for DCV*

Figure 2: *Sample page from the DCV questionnaire*

**Supplementary File 2**

*Self-efficacy items included in DCV task*

Please rate in each of the blanks in the column how certain you are that you can perform behaviour X

*Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:*

0	10	20	30	40	50	60	70	80	90	100
Cannot										Highly certain
do at all										can do

---

Please rate in each of the blanks in the column how certain you are that you could get yourself to perform behaviour X

*Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:*

0	10	20	30	40	50	60	70	80	90	100
Could not										Highly certain
do at all										could do

---

Running head: WHAT do SE items measure?

Please rate in each of the blanks in the column how certain you are that you could get yourself to perform behaviour X if you wanted to:

*Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:*

0      10      20      30      40      50      60      70      80      90      100

Even if I wanted to,

If I wanted to,

certain could not do at all

highly certain could do

---

I am confident that I can perform behaviour X

0            1    2    3    4    5    6    7    8    9    10

Does not

Not at all

Very

apply to me

Confident

Confident

---

I am confident that I could perform behaviour X

0            1    2    3    4    5    6    7    8    9    10

Does not

Not at all

Very

apply to me

Confident

Confident

(continued.) *Self-efficacy items included in DCV task*

---

I am confident that I can perform behaviour X if I wanted to





Strongly Disagree							Strongly Agree
I (BEHAVIOUR X) because I simply enjoy it.							
0	1	2	3	4	5		
Strongly Disagree							Strongly Agree

‘OTHER’ ITEMS

I am the kind of person who does (BEHAVIOUR X)							
0	1	2	3	4	5		
Strongly Disagree							Strongly Agree
I need to concentrate to do (BEHAVIOUR X)							
0	1	2	3	4	5		
Strongly Disagree							Strongly Agree

OPPORTUNITY ITEMS

There are many places where I can do (BEHAVIOUR X) at home and at work							
0	1	2	3	4	5		
Strongly Disagree							Strongly Agree
Most of the people I know do (BEHAVIOUR X) often							

Running head: WHAT do SE items measure?

0	1	2	3	4	5
Strongly Disagree					Strongly Agree

#### OUTCOME EXPECTANCY ITEMS

(BEHAVIOUR X) would help me reduce tension and manage stress					
0	1	2	3	4	5
Not at all					Very much
(BEHAVIOUR X) would help me control my weight					
0	1	2	3	4	5
Not at all					Very much