

Validation and Validity: Self-Efficacy, **Metacognition, and Knowledge-and-Appraisal Personality Architecture (KAPA)**

Laura Ritchie, Benjamin T. Sharpe

College Lane, University of Chichester, West Sussex, UK Email: l.ritchie@chi.ac.uk, b.sharpe@chi.ac.uk

How to cite this paper: Ritchie, L., & Sharpe, B. T. (2022). Validation and Validity: Self-Efficacy, Metacognition, and Knowledge-and-Appraisal Personality Architecture (KAPA). Psychology, 13, 1474-1503. https://doi.org/10.4236/psych.2022.1310094

Received: June 19, 2022 Accepted: September 27, 2022 Published: September 30, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

 $(\mathbf{\hat{n}})$ **Open Access**

Abstract

This research examines and reframes the long-standing pursuit of validation in personality research, in response to a shift in the field to focus on inter-individual idiographic contexts and constructs by focusing on Knowledge-and-Appraisals Personality Architecture (KAPA) as a method for person-centred, non-trait-based personality research. Aspects of instrument validity and validation are addressed, exploring relationships of cognition with self-variables to support the unique and informative nature of KAPA model for measuring personality architecture. First the validity of two empirical questionnaires, Self-efficacy for Performing (SEP) and the Metacognitive Thinking Questionnaire (MTQ) are demonstrated through EFA and CFA. Then the KAPA model for personality architecture, a person-based method that sits outside the traditional questionnaire approach to measuring personality, is validated by using elements of traditional validation and through exploring theoretically expected ontological relationships.

Keywords

Metacognition, Self-Efficacy, Personality, Validation, KAPA, Personality Measurement, Questionnaire Development

1. Introduction

Innovation and theoretical advances in research can take decades to adopt (Yu & Hang, 2010; Petzold et al., 2019), and changes of approach in personality research are no exception. William James advocated a person-centred approach with aspects of cognition being central to functioning, yet history adopted behaviourism. Rogers (1940) brought the "person" to the forefront in the 1960s and Bandura's (1986) social cognitive theory developed from his social learning theory (Bandura, 1969). Now, decades later, social cognitive theory is widely accepted as integral to and underpinning human functioning.

Since Allport and Odbert's (1936) original categorisation of words to "distinguish the behaviour of one human being from another" (p. 24), personality research has experienced theoretical postulates pushing beyond the concept of traits. Sullivan (1953) indicated the importance of intrapersonal situations and concepts; Mischel (1968) highlighted the contextual situation; Carlson (1971) questioned moving away from the "person"; and Mischel and Schoda (1995) introduced recognising individual differences. Bandura argued that personality is rooted in agency and not adequately measured by "behavioural clusters" represented by traits (Bandura, 1999: p. 23). More recently Cervone (2004) presented a method for measuring the architecture of personality, adopting person-specific methods and recognising idiosyncratic context-driven expressions of personality. Despite these advances, personality research continued to focus on less person-centred approaches by measuring pre-defined traits. Within the field theorists and practitioners continue to state the need to move from data devoid of inter and intrapersonal context (Borsboom, Mellenbergh, & Van Heerden, 2004; Pervin, 1994; Uher, 2013; Matthews, 2018) which forces the re-examination of methods and their validity to take new research forward.

1.1. Aims

The present research aims to address the validity of three psychological instruments: 1) Self-efficacy for Performing (SEP); 2) Metacognitive Thinking Questionnaire (MTQ); 3) Knowledge-and-Appraisals Personality Architecture (KAPA). The SEP and MTQ are new or adapted questionnaires and thus have not been validated, and the KAP has been used in research for decades, but there has been no explicit validation study on this method. Examining the first two questionnaires enables a discussion of the validity of the more elaborate KAPA methodology, and together these demonstrate the alignment of person-centred, nontrait based personality research with social cognitive theory and provide researchers with robust tools to use in future studies.

Aim 1 is to validate an adapted version of the Self-efficacy for Performing questionnaire (SEP), originally validated through EFA as a music-specific scale (Ritchie & Williamon, 2011). The SEP in the present research moves away from the original musical context, and then the original validation study is mirrored using EFA and continued by employing CFA with a discrete second sample to complete the validation process.

Aim 2, the development and validation the new Metacognitive Thinking Questionnaire (MTQ) is carried out following steps for internal and construct validity as outlined by Churchill (1979) including using two discrete samples to carry out EFA and then CFA. The new MTQ is then compared two other established metacognition questionnaires to demonstrate its robustness.

Finally, aim 3 the validity of the Knowledge-and-Appraisals Personality Ar-

chitecture (KAPA) is demonstrated using the full sample of participants. The 30 internal items of the KAPA model presented in this research have been constructed as a possible base for future versions of the KAPA model. Internal reliability of the present model and construct validity of the KAPA model approach to personality architecture and measurement are explored through internal and external relationships with other constructs (self-efficacy and metacognition).

This study received ethical approval from the University of Chichester Research Ethics Committee, approval number 2021_29.

1.2. On Validation

Research has explored validation since the introduction of the standardised test (Sireci, 2020; Thorndike, 1904). The first formal methodological guidelines for testing validity, the *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (American Psychological Association et al., 1954), introduced the concept of construct validity which was further explained by Cronbach and Meehl (1955). This was the beginning of understanding that the relationship between the concept of a construct and its measurement was central to validation studies.

In subsequent publications AERA et al. (1974, 1985) separate facets of validity (e.g., content validity), explaining various aspects or types of validity could support the overall construct validity, which remained overarching. Messick, a prominent validity theorist, stated "the meaning of the measure, and hence its construct validity, must always be pursued—not only to support test interpretation, but also to justify test use" (Messick, 1989: p. 17). Most recently, American Educational Research Association et al. (2014) make clear that "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11), with sources from the test's content, response processes, internal structure, relationships to other variables, and the consequences of testing (see pp. 16-21).

The "reliability/precision" of an instrument (American Educational Research Association et al., 2014) and its interpretability are paramount. Sireci (2007) illustrates a challenge in demonstrating validity by contrasting the lack of translatability and understanding latent variables with the ease of understanding the concrete notion of content validity. It is unfortunately common practice, specifically in personality psychology, to focus on content validity without reporting beyond alpha coefficients (for a review of practice, see Flake et al. (2017). Validation concepts and practices continue to evolve as authors fill gaps in research and present an array of evidence for validity appropriate to the measures under investigation (Borsboom, Mellenbergh, & Van Heerden, 2004; AERA et al., 2014; Flake et al., 2017; Sireci, 2020).

1.3. Self-Efficacy

The accumulated research on self-efficacy, focusing on personal beliefs about capabilities for task delivery (Bandura, 1977, 1997) is vast; with over 56,000 in-

dexed studies on the PsychInfo database. Self-efficacy research has demonstrated its relevance across areas of human agency including goal setting (Zimmerman et al., 1992; Huang, 2016), thought processes and strategy use (Schunk & Gunn, 1986; Bandura, 1989; Phan, 2009), achievement (Locke et al., 1984; Schunk & Usher, 2011), and adaptability to behavioural change (Cervone, 2000; Short & Ross-Stewart, 2008). The interest in self-efficacy and its relevance cannot be understated.

Self-Efficacy Measurement

Self-efficacy is task-specific and criterially based, therefore the approaches to measurement have varied greatly, from Bandura's original presentation of practical tasks to questionnaires containing one or many items (Bandura, 1977; Berry et al., 1989; Nicholas et al., 2015). Some scales claim to measure "general" self-efficacy (Sherer et al., 1982; Luszczynska et al., 2005), which arguably address more global constructs like self-image or self-esteem, as opposed to specific self-beliefs to carry out a criterial task. Bespoke, purposefully devised scales measure specific tasks within domains (Ritchie & Williamon, 2011; Tsai et al., 2019). When the task as an integral part to the questionnaire items, this ensures adherence to the construct yet limits future use beyond that task setting (see advice on scale construction by Bandura (2006) and Bong (2006)).

The Self-efficacy for Performing scale in the present research was originally validated alongside the Self-efficacy for Learning scale, demonstrating distinct types of self-efficacy within music (Ritchie & Williamon, 2011). These aimed to adhere to self-efficacy theory retain usefulness for future research. An adaptable preamble introduced the specific task and scale items described facets of skill de-livery instead of focusing on task minutiae. Validity was demonstrated through internal reliability, EFA, and through correlations with specific musical skills and attributes, and test-retest reliability. However to date no study has conducted a confirmatory factor analysis of this scale. The SEP questionnaire was adapted for sport and demonstrated the same internal reliability and factor structure with EFA (Ritchie & Williamon, 2012).

1.4. Metacognition

Bandura's (1986) concept of thought mediating action is made manifest in metacognition. Flavell (1979) defined metacognition as involving monitoring and controlling cognition and outlined the involvement and importance of cognitive self-appraisal, which includes aspects of knowledge about the person, task, and strategy. For Flavell, "person knowledge" "encompasses everything that you could come to believe about the nature of yourself and other people as cognitive processors" (p. 907). This aligns with Bandura's (1986) social cognitive theory, where personal, environmental, and behavioural elements connect throughout human functioning, on the micro-level. Interestingly, Flavell's (1979) definition of the person category aligns with tenants of self-efficacy, encompassing self-beliefs about capabilities to carry out a task—using Flavell's terminology—as a cognitive processor. Bandura (1977) introduced self-efficacy two years earlier, and although not explicitly, Flavell describes metacognitive thoughts with very similar wording such as: "you may feel that you are liable to fail in some upcoming enterprise" (Flavell, 1979: p. 908).

Metacognition is not abstract, un-situated or decontextualised thinking; an implied "task" is associated with "thinking about thinking" (Wellman, 1985). Thinking suggests active engagement *with* something. Brown (1987) distinguishes between knowledge and regulation within metacognition. Knowledge involved conscious reflection on a task's requirements, and regulation involved implementation of self-regulatory strategies to accomplish the task.

Following Flavell, subsequent researchers expanded, clarified, and qualified aspects of metacognition while contextualising metacognition to tasks within domains. Vandergrift et al. (2006: p. 435) state "metacognition is both self-reflection and self-direction" when learning a second language. Jayapraba (2013) highlight the importance of "ordered processes used to control one's own cognitive activities and to ensure that a cognitive goal has been met" (pp. 165-166) within science teaching. Pearman et al. (2020), stress the importance of actively engaging with self-reflection, to avoid unconsidered, habitual responses in ageing populations.

Metacognition Measurement

Numerous measurement tools and interventions have measured metacognition in diverse areas including health and treatment contexts (Clark et al., 2003; Bailey & Wells, 2015) and varied academic disciplines (see Veenman et al., 2006; Hacker et al., 2009; McCormick et al., 2012 for reviews of educational studies and practices; see also Kinnunen & Vauras, 1995; De Jager et al., 2005; Zohar & Barzilai, 2013; Garrison & Aykol, 2015). Pintrich and de Groot (1990) devised the Motivational Strategies for Learning Questionnaire (MSLQ) for educational contexts which addresses metacognition but focuses on self-regulated learning, which is not uncommon in educational studies (see Zimmerman & Schunk, 2001).

The Metacognitions Questionnaire 30 (MCQ-30) (Wells & Cartwright-Hatton, 2004) has been widely used and translated into multiple languages. Although it was designed to measure metacognition, subscales do not all consider "thinking about thinking". For example, one subscale focuses on negative aspects of memory with statements including "I do not trust my memory" and "I have a poor memory". The Vandergrift et al. (2006) metacognitive awareness listening questionnaire is theoretically sound, yet it contains items so task-specific to learning a second language that it is not useful for researchers outside this context.

Schellings and Van Hout-Wolters (2011) discuss self-report measures and suggest careful consideration by researchers of how tasks generalise, allowing for future usefulness of questionnaires, and they also advise a mixed method approach. Uniquely, Van Gog & Scheiter (2010) used eye tracking to supplement self-reports.

1.5. Personality

Personality research has an extended history which has been dominated by ex-

tremely popularised instruments such as the Myers-Briggs Type Indicator (MBTI) (Briggs & Myers, 1977) which roots in Jung's (1921) personality types and variations on the "Big-Five Structure" (Goldberg, 1993; John et al., 1991). The Big-Five has been validated across domains and cultures (Denissen et al., 2008; Kleinstäuber et al., 2018; Kohút et al., 2021). Despite this popularity, theoreticians urge a shift from measuring traits to measuring persons (Carlson, 1971; Mischel, 1973; Molenaar, 2004; Cervone, 2005; Beckmann & Wood, 2017; Renner et al., 2020).

Cervone's (2004, 2021) KAPA method of personality architecture uniquely differs from trait-based approaches to measuring personality. KAPA focuses on within-persons analysis without presupposing participants to "be" within any pre-defined category.

Personality Measurement with KAPA

Within KAPA, individuals identify their own strengths and weaknesses, contextualise these within relevant life-situations, and then rate the likelihood they would successfully undertake each of these situations. Cervone's (2004) original KAPA research utilised 83 contextual situations representing aspects of life experienced by an undergraduate population. These 83 items were not intended either to be definitively used in future studies, but were designed for population. KAPA is by design malleable. Subsequent research used the KAPA model across domains, with varying numbers of tailored items, ensuring relevance to practical life experiences. Studies have included academic settings of psychology (Calarco et al., 2015) business studies (Artistico & Rothenberg, 2013), and physical recreational (Wise, 2007), the workplace (Hoffner, 2006, 2009), clinical psychotherapy (Scott et al., 2021), and rehabilitation settings (Cervone et al., 2008).

The validity of KAPA has not been formally investigated, perhaps because it is not a simple questionnaire. KAPA includes clear methodological processes, yet its sub-items are malleable. Most recently, McKenna et al. (2021) devised 30 ideographically tailored items. Because of this variability, KAPA cannot be validated strictly as "a" definitive instrument and traditional empirical validation would be inappropriate. However, Borsboom, Mellenbergh and Van Heerden (2004) suggest demonstrating validity through ontological relationships. McKenna et al. (2021) discuss the theoretical underpinning of KAPA, specifically considering the relevance of measuring person-centred personality, instead of imposing an external structure. They facilitated that discussion by designing 30 items loosely aligned with Big-Five factors. For the present research, this loose association allows a comparative discussion demonstrating construct validity of KAPA.

2. Materials¹

All participants completed the complete battery of questionnaires including questionnaires on self-efficacy, metacognition, and the KAPA model of personality architecture. The Self-efficacy for Performing questionnaire (SEP), origi-

¹The instruments validated in this research appear in the Supplemental Materials.

nally validated by Ritchie and Williamon (2011) comprises 9 items which yield one summative self-efficacy score (min 9, max 63). Minimal wording adaptations were undertaken to remove any musical references: replacing "the performance" with "the task" and "playing", and "the music" became "skills" and "the task". A preface asking participants to consider a specific task while completing the questionnaire was retained. Participants also named their profession, described the main task in their profession, provided a single numeric representation of their self-efficacy to carry out that task on a 100-point scale, and provided a free-text typed explanation describing their confidence for this task.

The Metacognitive Thinking Questionnaire (MTQ) was closely aligned to Flavell's (1979) definition of metacognition. The understanding of "processes" was central to the new MTQ; it aimed to reflect alignment of social cognitive theory with metacognition, specifically the individual's capability to have, direct, and regulate thoughts.

The MTQ uses a 7-point Likert-type scale based on the widely used Garrison and Akyol's (2015) 6-point "metacognitive construct for communities of individuals" (henceforth referred to as GA13) with labels ranging from "very true of me" to "very untrue of me".

The GA13 and the Wells and Cartwright-Hatton (2004) metacognitions questionnaire (MCQ-30) were examined as established, validated questionnaires to compare validity and efficacy of the MTQ to other questionnaires in the field. The version of the GA13 to assess individuals (there is also a group version) was included in the present research. Internal scale items represent knowledge, monitoring, and regulation of cognition. The Wells and Cartwright-Hatton (2004) MCQ-30 is also widely used to study metacognition, and comprises five subscales covering cognitive confidence, positive beliefs, cognitive self-consciousness, uncontrollability and danger, and the need to control thoughts. The 30 items are rated on a 4-point Likert-type scale. Subscales focus strongly on a single topic, such as "memory" or "worry", as opposed to focusing on strategic processes surrounding different facets of cognition.

KAPA (Cervone, 2004, 2021) maps personality architecture through multiple components requiring self-reported, free-text appraisals and descriptions of personal strengths/weakness, a sorting task rating the relevance of the strength/ weakness in relation to contextualised settings, and numerical ratings of the likelihood to succeed in the same contextualised situations (representing self-efficacy). The 30 internal items from McKenna et al. (2021) were adapted through minimal wording changes removing the student-specific academic context references to enable applicability to a wider population. For example, "assignments" became "things" and "parents" became "relatives".

3. Method

3.1. Participants

228 participants aged between 18 - 77 were recruited via online networks, in-

cluding sending emails through university systems and social media (Facebook, Twitter, and Mastodon), and completed online questionnaires (compiled via Qualtrics software), commencing with an information sheet and consent form. Initial questions covered demographic information, including age and gender alignment (on an 11-point scale with the option of "I do not align with this scale"). The stand-alone self-efficacy metric was collected and then participants completed the KAPA, SEP, MTQ, GA13, and MCQ-30. The validity of the SEP and MTQ is considered by dividing the sample into two discrete groups, one sample of 50 with 10 males, 38 females; two undeclared/non-binary, and the second sample of 178 with 65 males, 83 females, and 30 undeclared/non-binary.

All who fulfilled the criteria of being 18 or over and signing the consent form and completed the questionnaire in full were included in the analysis. Rationale for dividing the sample is explained below in section 3.2.

3.2. Planned Analyses

The analysis of the SEP and MTQ follow the validation suggestions outlined by Churchill (1979). Kasier-Meyer-Olkin Measure of Sampling Adequacy (KMO) and Bartlett's Test of Sphericity (with a minimum significance level of .05) were used to test the correlations between scale items in the MTQ prior to employing EFA with the initial sample. A minimum level of $\alpha = .7$ was considered acceptable for Cronbach's Alpha (Cortina, 1993; Kline, 1999).

The recommendations for an acceptable sample size for EFA range from a suggested acceptable participant to item ratio as low as 3:1 (Cattell, 1978) to using a sample size with a ratio of at least 10:1 (Everitt, 1975). Others have suggested that 50 is an adequate minimum sample size (Gorsuch, 1974; Velicer & Fava, 1998). Geweke and Singleton (1980) tested samples as small as 10 and concluded that 30 was adequate and when communalities were high and the number of factors was small, sample sizes below 50 were shown to be reliable by using Monte Carlo analyses (Mundfrom et al., 2005). De Winter et al. (2009) undertook extensive Monte Carlo simulations to demonstrate the success of factor analysis with varying numbers of factors, internal items per factor, and levels of communalities and for a single underlying factor, they found "factor recovery can be reliable with sample sizes well below 50" (p. 153). In the present research both scales to be tested have one hypothesised factor with several items to load onto it, and the minimum ratio of 3:1 was considered when deciding on a sample of 50. We also adopted a rule of thumb where sample size for CFA is a minimum of the number of measurement variables (items in questionnaire) * 10. Green's (1991) rule of thumb (medium effect) was considered when calculating sample size requirements for regression analysis (N \ge 50 + (8 * No. Variables). With 16 items in the MTQ and 9 in SEP, 178 participants were needed to carry out CFA.

As opposed to only using chi-square likelihood ratio test statistics, multiple measures of fit indices were considered to provide a more accurate model evaluation process (Byrne, 1998; Hoyle, 1995; Kline, 1998; Tanaka, 1993). However,

only appropriate measures are reported. As both SEP and MTQ do have correlated component items, and are not testing a null hypothesis, some indices are not relevant to the validation of these questionnaires.

CMIN is not reported, since in Amos this is the chi-square value. Chi-square has notable problems that occur when the sample size exceeds 200 (Alavi et al., 2020), and the Satorra-Bentler scaled chi-square which addresses some of these issues was deemed preferable and therefore reported. The NFI tests the null model where components are uncorrelated (Byrne, 1994), and was therefore not appropriate here. PCFI (Blunch, 2008) is a parsimony-corrected index and without an overly complex model, this measure is inappropriate and unnecessary. As we do not have a null model, GFI will not be reported (Hu & Bentler, 1995). We instead use incremental indices (Hu & Bentler, 1995, 1999) which "measure the appropriateness of fit of a hypothesised model compared with a more restricted, albeit nested, baseline model" (Byrne, 2013: p. 70). CFI (>0.95; Hu & Bentler, 1999) and RMSEA (90% CI, <0.05; Browne & Cudeck, 1993) are reported. RMSEA shows how well the hypothesised model fit the sample data; it is sensitive to model misspecification (Hu & Bentler, 1998); and it is possible to build confidence intervals around this statistic. AIC values are reported (Lower = better; Raykov & Marcoulides, 2000). Construct validity was also demonstrated by relationships with other constructs; Pearson Correlations were carried out with all variables.

To identify whether self-efficacy and metacognition uniquely predict behaviours scoring high (more likely) in relation to the self-declared personality strength/weakness in KAPA, multiple linear regression analysis was conducted. Beta coefficients (β) were used to access the unique variance associated with each variable.

4. Results

4.1. Self-Efficacy for Performing (SEP)

The Kasier-Meyer-Olkin Measure of Sampling Adequacy (KMO) coefficient was .742 (above the suggested level of .6) and Bartlett's Test of Sphericity was highly significant, $x^2(36) = 183$, p < .001. The sample demonstrated a range of self-efficacy scores, from 27 to 63 (M = 44.5). The SEP yielded good internal reliability, with $\alpha = .826$. EFA using parallel analysis with Maximum Likelihood extraction and Quartimax rotation, as suggested when a single underlying factor is hypothesised (Stewart, 1981; Gorsuch, 1983), replicated the original validation results of Ritchie and Williamon (2011), with a single underlying factor, and a shadow factor representing reverse-coded items (also seen in Gaudry et al., 1975). (See Table 1) The consistency with previous published results demonstrated this adapted SEP maintained efficacy thus far in the validation process.

CFA

CFA was conducted using AMOS statistics 23. Models with one and two factors (to include the shadow factor) were explored. The model with the shadow factor

produced a moderate fit. After examining residuals in the correlation matrix to gain meaningful information regarding the CFA model (Byrne, 2010), errors of two internal items were found to be correlated. (See Figure 1) Following Jackson et al. (2009) correlating errors of items is acceptable when they contain related words or phrases.

The resulting model produces a good fit: (Satorra-Bentler $\chi^2(25) = 35.867$, p = .074, CFI = .975, RMSEA = .050 [.000, .084]), Model AIC = 75.867. Theoretically, both the main and shadow factors represent self-efficacy, as opposed to representing a second construct. The shadow factor presents a response effect (DiStefano & Motl, 2006). Borgers et al. (2004) also found this response effect and confirmed that it had no impact on reliability measures. This CFA model demonstrates a good fit, with robust factor loadings and strong expected relationships between the two factors of positive and negatively worded items.

	Factor 1	Factor 2
SEP1	571	.348
SEP2	.722	
SEP3	.585	
SEP4	.725	
SEP5		.838
SEP6	.775	
SEP7	.557	
SEP8	.739	
SEP9	328	.690

Table 1. EFA factor loadings for the SEP*.

*Note: Minimum Residuals extraction; Quartimax rotation.





4.2. Metacognitive Thinking Questionnaire (MTQ)

The Kaiser-Meyer-Olkin coefficient for the MTQ scale was .832, indicating the sample to be adequate, and Bartlett's test for sphericity was highly significant, $x^2(28) = 223$, p < .001, confirming the data is acceptable for EFA. Internal reliability was also tested, and the scale showed an unacceptably high Cronbach alpha, a = .934, beyond the recommended level of acceptability of .90 (Streiner, 2003), suggesting several internal items may measure the same thing. Therefore, the MTQ items needed to be examined either statistically or theoretically to consider removing some of the items. The internal structure was explored with EFA, using parallel analysis, and following the planned analysis using the Varimax orthogonal rotation as suggested by Kline (1998), and results showed a single underlying factor which did not provide statistical insight for the removal of items.

The 16 items were theoretically examined and a clear division emerged between those explicitly mentioning a task and generalised, or purely conceptual items. Metacognition is an applied and developed process encompassing aspects of self-regulatory strategies, self-beliefs, and aligning with self-efficacy by encompassing an awareness of a person's thoughts concerning their agentic potential toward a goal. Generalised items lacked specificity, and diluted the usefulness of a questionnaire measuring metacognitive thinking as a directed activity *about* (a task) (Livingston, 2003; Veenman et al., 2006). Eight items were retained in the MTQ, and produced an acceptable $\alpha = .898$. EFA showed one factor. (See **Table 2**)

4.2.1. CFA

The MTQ's internal reliability was confirmed with data from a second independent sample of 178 (α = .883). CFA, using AMOS statistics 23, was conducted specifying a single factor model, following the results of the EFA. Results were examined to gain meaningful information regarding the model (Byrne, 2010), and errors from two pairs of internal items were shown to be correlated. The correlated pairs use similar words and phrases and thus it was both theoretically sound and acceptable to include this modification in the CFA model (Jackson et al., 2009). The CFA model, shown in Figure 2, resulted in a very good fit with all coefficients within acceptable levels: Satorra-Bentler $\chi^2(18) = 23.549$, p = .170., CFI = .991, RMSEA = .042 [.000, .084], Model AIC = 59.549.

4.2.2. Construct Validity

Reliability and structural validity of the MTQ was compared to that of other two established metacognition questionnaires, the Garrison and Akyol (2015) GA13 and the Wells and Cartwright-Hatton (2004) MCQ. The GA13 produced a slightly higher than acceptable alpha (α = .903). When the MCQ-30 was tested for internal reliability as a 30-item instrument, it appeared to produce an acceptable Cronbach alpha of .841. However, the MCQ-30 comprises five subscales representing separate factors. When testing subscales for internal reliability, results varied from the low side of acceptability for "cognitive confidence" (α = .633), to overly high alpha for "cognitive self-consciousness" (α = .915) and

"uncontrollability and danger" (α = .912), however the "need to control thoughts" and "positive beliefs" both produced acceptable alpha (α = .787, α = .893).

Relationships between the metacognition scales and self-efficacy beliefs were examined to demonstrate construct validity and robustness of the measures. Data from all questionnaires satisfied tests for normality (George & Mallery, 2010). Summative scores were created to allow subsequent comparative analyses: MTQ (M = 42.8, SD = 7.75); GA13 (M = 61.1, SD = 9.97); MCQ-30 (M = 70.7, SD = 12.6). There was an expected relationship with self-efficacy, which encompasses aspects of cognitive thought directed toward task delivery. As shown in **Table 3**, the MTQ consistently correlated highly with self-efficacy scores from the SEP and the numerical rating of self-efficacy to carry out the main task of people's profession. The MTQ produced more significant relationships than the GA13 or MCQ-30. The theoretical underpinning of the MCQ-30 scale was not in alignment with the definition of metacognition presented by Flavell (1979), and did not focus on activities directed toward a task. This dissonance was reflected in the lack of correlations produced by the MCQ-30 with self-efficacy.

	Factor1
MTQ5	.853
MTQ7	.667
MTQ8	.714
MTQ11	.764
MTQ12	.727
MTQ13	.637
MTQ14	.723
MTQ16	.705

Table 2. EFA factor loadings for the MTQ*.

*Note: Maximum Likelihood extraction; Varimax rotation.



Figure 2. CFA model for the metacognitive thinking questionnaire.

	MTQ	GA13	MCQ-30
SEP	.429***	.352***	101
SE Job	.390***	.304***	080

Table 3. Correlations between the MTQ, GA13, MCQ-30 and self-efficacy for performing (SEP) and self-efficacy for completing the main task of one's profession (SE Job)*.

*Note: *** *p* < .001.

To compare the CFA results of the MTQ, CFA was also undertaken with the GA13 and the MCT30, which are presented in the literature as validated (Wells & Cartwright-Hatton, 2004; Garrison & Akyol, 2015). Neither the GA13 nor the MCQ-30 produced acceptable CFA results. The GA13 was tested with one factor, as indicated by Garrison and Akyol (2015), and produced: Satorra-Bentler χ^2 (65) = 533.876, *p* < .000., CFI = .761, RMSEA = .202 [.186, .218], Model AIC = 585.876. CFA was also undertaken with the MCT30 using its intended five factor structure, and none of the resulting coefficients were within acceptable permeameters: Satorra-Bentler χ^2 (405) = 717.267, *p* = .000., CFI = .885, RMSEA = .066 [.058, .074], Model AIC = 837.267. For reference, the Satorra-Bentler χ^2 should have a non-significant *p* value which is >.05, CFI represents a good fit when it is greater than 0.95 (Hu & Bentler, 1999); RMSEA values closer to 0 represent a good fit (with a 90% confidence interval, and *p* < .05; Browne & Cudeck, 1993); and lower values for Model AIC coefficients are considered better (Raykov & Marcoulides, 2000).

The MTQ demonstrated a robust internal structure, satisfied all the fit indices as reported above, and when compared to the other metacognition measures it demonstrated the strongest and most significant correlations with self-efficacy (both SEP and SE Job). Further relationships between the MTQ and aspects of personality architecture are subsequently explored with the KAPA model.

4.3. KAPA Model

Responses from the 228 participants, representing a full data set, were considered. The internal items of the KAPA inventory used in this research were not being validated as comprising a definitive instrument, however some traditional validation tests have been performed to demonstrate integrity where appropriate and applicable, and to illustrate why traditional tests are not always appropriate. For example, the 30 sub-items used in the KAPA model in this research, representing life-situations and contexts, were tested for internal reliability and demonstrated a robust a Cronbach alpha score of .810. KAPA inherently focuses on the individual, and thus the internal items that any researcher uses must be relevant to their specific population. Undertaking basic internal reliability testing with any sample can give valuable information about spurious or redundant scale items. The sub-items of KAPA present contextual situations; the 30 in this study can act as a starting place for researchers to further develop their own items as relevant to their research contexts. Following the practices set out by Cervone (2004, 2021), participant responses to the 30 items were categorised by those rated most/least relevant to the reported strength/weakness. Negatively coded items were realigned and scores for the number of these most relevant items were created to allow for comparison and individual personality mapping. A mean self-efficacy score (confidence to engage with situations presented in the 30 items) was calculated for the items which fell into these most relevant categories with relation to the strength/ weakness. (See Table 4)

Descriptive statistics for SEP, participant self-efficacy ratings of their confidence to carry out the main task at their job (SE Job), and MTQ are presented for the full sample of 228, along with the mean number of KAPA sub-items strongly related to participant strengths and weaknesses in the supplemental material. A paired sample t-test confirmed the difference in means visible between the buoyant self-efficacy for strengths (SE Strengths) and the low self-efficacy for situations where the weakness is least relevant (SE Weaknesses): t(217) = 25.6, p < .001.

KAPA results were compared with external constructs to demonstrate aspects of validity. The 30 sub-items were originally created by McKenna et al. (2021) to have a loose relationship to the factors of the Big 5 (Goldberg, 1993) inventory to enable comparative discussion about the approach to personality measurement. In the present study the loose association with the five personality categories allows for an illustration of why simply using traditional empirical validity tests (EFA/CFA) does not work with KAPA. When the 30 sub-items in the KAPA inventory are treated as factors (in line with the traits of the five-factor model) results produce some relationships with external constructs, as can be seen in Table 4, however, these limited results both mask and lose the personalisation inherent in the KAPA model by inadequately and incorrectly representing the measurements obtained. KAPA demonstrates levels of individual granularity between the relationship of each sub-item, the participant's strength and weakness, and the wider interaction with other constructs. This detail is precisely what creates the *personal* architecture of KAPA that is so refreshingly unique and relevant to understanding an individual.

Table 4. Extrapolated summative scores from KAPA, based on the five factors: Neuron	otic-
ism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness,	and
their relationships to self-efficacy (SEP and SE Job) and metacognition (MTQ)*.	

	SEP	MTQ	SE Job
KAPA Neuroticism	.340***	.239***	.253***
KAPA Extraversion			
KAPA Open to Experience		.272***	
KAPA Agreeableness	.148*	.179**	
KAPA Conscientiousness	.267***	.352***	.198**

Note*: ** *p* < .01, * *p* < .001.

Individual responses demonstrate distinct personal differences. **Figure 3**, below, illustrates two people who both have the same strength, being determined, and have the same score for their self-efficacy to carry out the main task in their job (100% confidence). Yet, their scores representing self-efficacy to succeed in the 30 contextual situations presented in the KAPA inventory vary dramatically.

The level of personal idiosyncrasy revealed through KAPA becomes further complicated when examining how individuals mapped their strengths and weaknesses onto the 30 situations. For each person a unique set of situations are most relevant to their strengths and weaknesses. Even when people have the same overall score for self-efficacy, the same personal strength, and the same profession there are distinct differences in their personality architecture. For example one may be outgoing with strangers whereas another maintains interest in long arguments. The differences in personality architecture of three teachers who shared the personal strength of being determined, and had the same self-efficacy score is presented in the supplemental material.

Examining relationships between the results produced by KAPA and other constructs also supports its validity. The relevance of participant's strengths correlated to their self-efficacy to succeed all correlated with a significance level of p < .001 for each of the 30 items in KAPA. The full table can be seen in the supplemental material. The mean self-efficacy scores for the items for which the strength was most relevant and for which the weakness was most relevant correlated significantly with scores from the Self-efficacy for Performing questionnaire and Metacognitive Thinking Questionnaire. The mean self-efficacy scores for the most relevant situations to strengths (hereafter SE Strength) correlated to





SEP .293, p < .001 and to MTQ .240, p < .001, and the mean self-efficacy scores of the most relevant situations to weaknesses correlated to SEP .179, p < .01, and MTQ .200, p < .01. The positive relationship between the constructs examined is demonstrated through the covariance matrix in Table 5. These relationships are in line with ontological expectations.

A linear multiple regression analysis was conducted to determine the extent to which SE Strengths (representing the mean self-efficacy scores for the KAPA situations where the strength was most relevant) was significantly predicted by SEP, and MTQ, with 10.1% of the variance in the SEP score explained, adj. $R^2 = .093$, F(2, 224) = 12.5, p < .001. (See Table 6)

5. General Discussion

This research demonstrates the need for different considerations necessary to satisfy technical requirements of validation, including the importance of considering the theoretical relevance and alignment of tools to the constructs they measure.

CFA was undertaken with the adapted version of the SEP questionnaire to demonstrate its validity. Previous research explored EFA, test-retest reliability, and construct validity through external relationships with the musical version of the SEP (Ritchie & Williamon, 2011), however no previous research had confirmed the scale through CFA. The scale produced consistently robust alpha scores, showing internal reliability, and patterns demonstrated with the EFA replicated previous results. The model for CFA required theoretical consideration and an examination of internal relationships of the positive and negatively worded items allowed for a model that satisfied acceptable levels for model fit coefficients. When used in future studies, researchers should still undertake tests to make sure the scale is appropriate and robust for the population under investigation.

Table 5. Covariance of SE Strengths, SEP, and MTQ.

	SE Strengths	MTQ	SEP
SE Strengths	1.830		
MTQ	2.940	66.650	
SEP	3.148	29.744	76.962

Table 6. Linear Multiple Regression predicting SE Strengths by SEP and MTQ.

	β	95% CI for β		95% CI for β		95% CI for β		В	р	<i>r</i> ² a(b, c)
		LL	UL							
SEP	.189	.052	.326	.029	.007	.030				
MTQ	.189	.052	.326	.031	.007	.030				

The MTQ was a new questionnaire, and the internal items of the questionnaire had not been previously tested. Measuring internal reliability highlighted the possibility of redundant items, and this was addressed by examining the theoretical alignment of each item to the construct they were measuring: metacognition. In this case the reference to a specific task was a requirement for item inclusion and generic, decontextualised items were discarded from the questionnaire. This process serves as a reminder that statistical tests cannot serve as a blanket license that supersedes theoretical relevance to grant a questionnaire's validity or suitability for research.

This point was then exemplified by examining the results from other two published metacognition questionnaires. The internal items of the MCO-30 initially demonstrated a seemingly robust alpha, but when considered within their intended sub-scales, results were much more variable. The GA13 and the MTQ demonstrated both internal reliability and expected relationships with the selfefficacy scores collected. Within psychology studies, alpha is the most commonly reported statistic to justify the acceptability of a scale for inclusion in research, and when examining the component items of these scales, it can be seen that the items contained in the GA13 and MTQ are more closely aligned to metacognitive theory, however neither of these satisfied tests for the CFA model fit. Several of the subscales of the MCQ-30 only tangentially reflect a current theoretical understanding of metacognition, and any measurements resulting from its use in research would likely misrepresent the construct. This represents a reminder and warning for researchers that examining both the integrity of an instrument's content and its relation to the construct are necessary (see also Hoekstra et al., 2019).

The validity of the KAPA model addresses the complex and ongoing quest for validity in personality research and specifically highlighted individual differences in personality architecture within a domain. Exploring the KAPA model for validity necessitated a different approach to empirical questionnaire validation. KAPA not a definitive instrument that could be "validated" through the same processes as the previous questionnaires, yet the infrastructure and methodology did need to undergo controlled testing to demonstrate both reliability and ontological relationships. The 30 items used with KAPA in demonstrated a robust internal reliability, and there were expected relationships with constructs (between SE Strengths and externally measured constructs (SE and MTQ)), and expectedly lacking relationships provide a new perspective which opens a door for the practical application of KAPA beyond a personal, diagnostic tool.

The KAPA data also illustrated multiple differences between people with the same overall "scores", demonstrating the level of personality detail captured through this method and how simple "trait" factors are inadequate to explain individuals at this granular level. Regression analysis demonstrated the predictive power of these constructs, which spoke to the relationship of these constructs to the tenants of social cognitive theory. Bandura (1986) outlined this

triadic reciprocity between personal, behavioural, and environmental components found in everyday life, and it occurs both on macro and micro levels. This reciprocity was demonstrated with the influence of self-efficacy, metacognition, and the interaction of self-efficacy, personal strengths, and life situations on one another.

6. Limitations

The processes undertaken in these studies to demonstrate validity are neither exhaustive nor final. More validation explorations will need to be undertaken to maintain the currency and relevance of these results with different samples, as cultures change, and as understandings of the constructs under investigation develop. Only three measures were considered here and there could be a contextualised real-life demonstration for these constructs to be more meaningful and relevant. Demonstrating the validity of measures is a very first step to prepare for future research. The present research represents a snapshot of these constructs and their relationships. Due to the person-centred focus of the KAPA model of personality architecture, wider generalisations of specific relationships are not possible. To illustrate further validity and test-retest reliability, future studies could undertake longitudinal or intervention studies with other variables.

7. Conclusion

This research built on the work of previous studies and considered advice on best and current practice within the field. The development of new scales continues to be relevant to research as understanding of constructs evolve, and responsibility falls on researchers to ensure their instruments suitable and ontologically substantiated. The examination of the three instruments in this research has demonstrated a malleable approach to validity which takes into consideration the type of instrument and its relationship to and with the constructs it measures and maps.

KAPA provides a practical means to identify the principal aspects of an individual's personality in relation to their own idiosyncratic context. In addition, the self-efficacy and metacognition tools can allow researchers to assess personal beliefs and thought mediated action. Together, these can contribute to the development of practical methods for determining an individual's awareness of their own capabilities and the level of cognitive engagement prioritised towards a given context. It is hoped that subsequent research using these tools moves the field of personality research towards a more person-centred approach.

Declarations of Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The studies presented here were registered on the Open Science Framework and

the full data set is available here: https://osf.io/y26mh/.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- AERA, American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1974). *Standards for Educational and Psychological Tests.* American Psychological Association.
- AERA, American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. American Psychological Association.
- AERA, American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014) *Standards for Educational and Psychological Testing.*
- Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-Square for Model Fit in Confirmatory Factor Analysis. *Journal of Advanced Nursing*, 76, 2209-2221. <u>https://doi.org/10.1111/jan.14399</u>
- Allport, G. W., & Odbert, H. S. (1936). Trait-Names: A Psycho-Lexical Study. Psychological Monographs, 47, 1-171. <u>https://doi.org/10.1037/h0093360</u>
- American Psychological Association, American Educational Research Association, & National Council on Measurements Used in Education (1954). Technical Recommendations for Psychological Tests and Diagnostic Techniques. *Psychological Bulletin, 51(2, Pt 2)*, 1-38.
- Artistico, R., & Rothenberg, A. M. (2013). Assessing Strengths and Weaknesses in Solving Work Problems: A Knowledge and Appraisal Personality Architecture (KAPA) Analysis of the Trait Conscientiousness and Self-Efficacy. *International Journal of Psychological Studies*, 5, 84-97. <u>https://doi.org/10.5539/ijps.v5n3p84</u>
- Bailey, R., & Wells, A. (2015). Development and Initial Validation of a Measure of Metacognitive Beliefs in Health Anxiety: The MCQ-HA. *Psychiatry Research, 230*, 871-877. <u>https://doi.org/10.1016/j.psychres.2015.11.035</u>
- Bandura, A. (1969). Social-Learning Theory of Identificatory Processes. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 213-262). Rand McNally & Company.
- Bandura, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review, 84, 191-215. <u>https://doi.org/10.1037/0033-295X.84.2.191</u>
- Bandura, A. (1986). Bandura Social Foundations of Thought and Action. Englewood Cliffs.
- Bandura, A. (1989). Regulation of Cognitive Processes through Perceived Self-Efficacy. Developmental Psychology, 25, 729-735. <u>https://doi.org/10.1037/0012-1649.25.5.729</u>
- Bandura, A. (1997). Self-Efficacy: The Exercise of Control. Freeman.
- Bandura, A. (1999). A Social Cognitive Theory of Personality. In L. Pervin, & O. John (Eds.), *Handbook of Personality* (2nd ed., pp. 154-196). Guilford Publications. (Reprinted in D. Cervone & Y. Shoda (Eds.), *The Coherence of Personality*. Guilford Press)
- Bandura, A. (2006). Guide for Constructing Self-Efficacy Scales. In F. Pajares, & T. Urdan (Eds.), Self-Efficacy Beliefs of Adolescents (pp. 307-337). Information Age Publishing.

- Beckmann, N., & Wood, R. E. (2017). Dynamic Personality Science. Integrating Between-Person Stability and Within-Person Change. *Frontiers in Psychology, 8*, Article No. 1486. <u>https://doi.org/10.3389/978-2-88945-345-0</u>
- Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and Validity of the Memory Self-Efficacy Questionnaire. *Developmental Psychology*, 25, 701-713. <u>https://doi.org/10.1037/0012-1649.25.5.701</u>
- Blunch, N. J. (2008). The Measurement Model in SEM: Confirmatory Factor Analysis. In Introduction to Structural Equation Modelling Using SPSS and AMOS (pp. 127-156). Sage.
- Bong, M. (2006). Asking the Right Question. In F. Pajares, & T. Urdan (Eds.), Self-Efficacy Beliefs of Adolescents (pp. 287-305). Information Age Publishing.
- Borgers, N., Sikkel, D., & Hox, J. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality and Quantity*, 38, 17-33. <u>https://doi.org/10.1023/B:QUQU.0000013236.29205.a6</u>
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The Concept of Validity. *Psychological Review, 111*, 1061-1071. <u>https://doi.org/10.1037/0033-295X.111.4.1061</u>
- Briggs, K. C., & Myers, I. B. (1977). *The Myers-Briggs Type Indicator: Form G.* Consulting Psychologists Press.
- Brown, A. L. (1987). Knowing When, Where, and How to Remember. A Problem of Metacognition. In F. E. Weinert, & R. H. Kluwe (Eds.), Advances in Instructional Psychology (Vol. 1, pp. 77-165). Erlbaum.
- Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Accessing Model Fit. Sage.
- Byrne, B. M. (1994). Structural Equation Modeling with EQS and EQS/Windows. Sage.
- Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming.* Erlbaum.
- Byrne, B. M. (2010). *Structural Equation Modeling with Mplus. Basic Concepts, Applications, and Programming.* Taylor & Francis Group.
- Byrne, B. M. (2013). Structural Equation Modeling with Mplus. Basic Concepts, Applications, and Programming. Routledge. <u>https://doi.org/10.4324/9780203807644</u>
- Calarco, D., Ferrante, D., & Di Blas, L. (2015). La coerenza della personalità da una prospettiva socio-cognitiva: Uno studio sulla variabilità intraindividuale dell'autoefficacia percepita in situazioni sociali. *Giornale Italiano di Psicologia, 42,* 829-852.
- Carlson, R. (1971). Where Is the Person in Personality Research? *Psychological Bulletin,* 75, 203-219. <u>https://doi.org/10.1037/h0030469</u>
- Cattell, R. B. (1978). Fixing the Number of Factors: The Most Practicable Psychometric Procedures. In *The Scientific Use of Factor Analysis in Behavioral and Life Sciences* (pp. 72-91). Plenum Press. <u>https://doi.org/10.1007/978-1-4684-2262-7_5</u>
- Cervone, D. (2000). Thinking about Self-Efficacy. *Behavior Modification, 24*, 30-56. https://doi.org/10.1177/0145445500241002
- Cervone, D. (2004). The Architecture of Personality. *Studia Universitatis Babes-Bolyai-Psychologia-Paedagogia*, 49, 3-44. <u>https://doi.org/10.1037/0033-295X.111.1.183</u>
- Cervone, D. (2005). Personality Architecture: Within-Person Structures and Processes. *Annual Review of Psychology, 56*, 423-452. https://doi.org/10.1146/annurev.psych.56.091103.070133
- Cervone, D. (2021). Five Paths to Personality Coherence: Integrative Implications of the Knowledge-and-Appraisal Personality Architecture. *European Journal of Personality*,

36, 319-346. https://doi.org/10.1177/08902070211015599

- Cervone, D., Caldwell, T. L., Fiori, M., Orom, H., Shadel, W. G., Kassel, J. D., & Artistico, D. (2008). What Underlies Appraisals? Experimentally Testing a Knowledge-and-Appraisal Model of Personality Architecture among Smokers Contemplating High-Risk Situations. *Journal of Personality*, *76*, 929-968. https://doi.org/10.1111/j.1467-6494.2008.00510.x
- Churchill, G. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research, 16*, 64-73. https://doi.org/10.1177/002224377901600110
- Clark, D. A., Purdon, C., & Wang, A. (2003). The Meta-Cognitive Beliefs Questionnaire: Development of a Measure of Obsessional Beliefs. *Behaviour Research and Therapy*, 41, 655-669. <u>https://doi.org/10.1016/S0005-7967(02)00032-3</u>
- Cortina, J. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Research*, 78, 98-104. https://doi.org/10.1037/0021-9010.78.1.98
- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. Psychological Bulletin, 52, 281-302. <u>https://doi.org/10.1037/h0040957</u>
- De Jager, B., Jansen, M., & Reezigt, G. (2005). The Development of Metacognition in Primary School Learning Environments. School Effectiveness and School Improvement, 16, 179-196. <u>https://doi.org/10.1080/09243450500114181</u>
- De Winter, J. C., Dodou, D., & Wieringa, P. A. (2009). Exploratory Factor Analysis with Small Sample Sizes. *Multivariate Behavioral Research, 44*, 147-181. <u>https://doi.org/10.1080/00273170902794206</u>
- Denissen, J. J., Geenen, R., Van Aken, M. A., Gosling, S. D., & Potter, J. (2008). Development and Validation of a Dutch Translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90*, 152-157. <u>https://doi.org/10.1080/00223890701845229</u>
- DiStefano, C., & Motl, R. W. (2006). Further Investigating Method Effects Associated with Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling*, *13*, 440-464. <u>https://doi.org/10.1207/s15328007sem1303_6</u>
- Everitt, B. S. (1975). Multivariate Analysis: The Need for Data, and other Problems. British Journal of Psychiatry, 126, 237-240. <u>https://doi.org/10.1192/bjp.126.3.237</u>
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, 8, 370-378. <u>https://doi.org/10.1177/1948550617693063</u>
- Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist*, 34, 906-911. <u>https://doi.org/10.1037/0003-066X.34.10.906</u>
- Garrison, D. R., & Akyol, Z. (2015). Toward the Development of a Metacognition Construct for Communities of Inquiry. *The Internet and Higher Education, 24*, 66-71. <u>https://doi.org/10.1016/j.iheduc.2014.10.001</u>
- Gaudry, E., Vagg, P., & Spielberger, C. (1975). Validation of the State-Trait Distinction in Anxiety Research. *Multivariate Behavior Research*, 10, 331-341. <u>https://doi.org/10.1207/s15327906mbr1003_6</u>
- George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update (10th ed.). Pearson.
- Geweke, J. F., & Singleton, K. J. (1980). Interpreting the Likelihood Ratio Statistic in Factor Models When Sample Size Is Small. *Journal of the American Statistical Association*, 75, 133-137. <u>https://doi.org/10.1080/01621459.1980.10477442</u>

- Goldberg, L. R. (1993). The Structure of Phenotypic Personality Traits. American Psychologist, 48, 26-34. <u>https://doi.org/10.1037/0003-066X.48.1.26</u>
- Gorsuch, R. L. (1974). Factor Analysis (1st ed.). W. B. Saunders Company.
- Gorsuch, R. L. (1983). Factor Analysis (2nd ed.). Lawrence Erlbaum Associates.
- Green, S. B. (1991). How Many Subjects Does It Take to Do a Regression Analysis. Multivariate Behavioral Research, 26, 499-510. https://doi.org/10.1207/s15327906mbr2603 7
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.) (2009). *Handbook of Metacognition in Education*. Routledge. <u>https://doi.org/10.4324/9780203876428</u>
- Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruyen, P. M. (2019). An Empirical Analysis of Alleged Misunderstandings of Coefficient Alpha. *International Journal of Social Research Methodology*, 22, 351-364. <u>https://doi.org/10.1080/13645579.2018.1547523</u>
- Hoffner, R. (2006). The Architecture of Personality in the Context of Work. Unpublished Master's Thesis, Virginia Polytechnic Institute and State University. <u>http://hdl.handle.net/10919/34885</u>
- Hoffner, R. A. (2009). Measuring Personality in Context: Improving Predictive Accuracy in Selection Decision Making. Doctoral Dissertation, Virginia Tech. https://vtechworks.lib.vt.edu/handle/10919/37859
- Hoyle, R. H. (1995). *Structural Equation Modeling: Concepts, Issues, and Applications.* Sage.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating Model Fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 76-99). Sage.
- Hu, L. T., & Bentler, P. M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychological Methods, 3*, 424-453. https://doi.org/10.1037/1082-989X.3.4.424
- Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55. <u>https://doi.org/10.1080/10705519909540118</u>
- Huang, C. (2016). Achievement Goals and Self-Efficacy: A Meta-Analysis. *Educational Research Review*, 19, 119-137. <u>https://doi.org/10.1016/j.edurev.2016.07.002</u>
- Jackson, D., Gillaspy, J., & Purc-Stephenson, R. (2009). Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. *Psychological Methods*, 14, 6-23. <u>https://doi.org/10.1037/a0014694</u>
- Jayapraba, G. (2013). Metacognitive Instruction and Cooperative Learning-Strategies for Promoting Insightful Learning in Science. *International Journal on New Trends in Education and Their Implications, 4*, 165-172.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory: Versions 4a and 54. Institute of Personality and Social Research, University of California.
- Jung, C. G. ([1921] 1971). Psychological Types. Collected Works, Volume 6, Princeton University Press.
- Kinnunen, R., & Vauras, M. (1995). Comprehension Monitoring and the Level of Comprehension in High- and Low-Achieving Primary School Children's Reading. *Learning* and Instruction, 5, 143-165. <u>https://doi.org/10.1016/0959-4752(95)00009-R</u>
- Kleinstäuber, M., Weise, C., Andersson, G., & Probst, T. (2018). Personality Traits Predict and Moderate the Outcome of Internet-Based Cognitive Behavioural Therapy for Chronic Tinnitus. *International Journal of Audiology*, *57*, 538-544. <u>https://doi.org/10.1080/14992027.2018.1432902</u>

Kline, P. (1999). The Handbook of Psychological Testing (2nd ed.). Routledge.

- Kline, R. B. (1998). Structural Equation Modeling. Guilford.
- Kohút, M., Kohútová, V., & Halama, P. (2021). Big Five Predictors of Pandemic-Related Behavior and Emotions in the First and Second COVID-19 Pandemic Wave in Slovakia. *Personality and Individual Differences, 180*, Article ID: 110934. <u>https://doi.org/10.1016/j.paid.2021.110934</u>
- Livingston, J. A. (2003). *Metacognition: An Overview*. https://files.eric.ed.gov/fulltext/ED474273.pdf
- Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of Self-Efficacy, Goals, and Task Strategies on Task Performance. *Journal of Applied Psychology, 69,* 241-251. https://doi.org/10.1037/0021-9010.69.2.241
- Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The General Self-Efficacy Scale: Multicultural Validation Studies. *The Journal of Psychology, 139*, 439-457. https://doi.org/10.3200/JRLP.139.5.439-457
- Matthews, G. (2018). Cognitive-Adaptive Trait Theory: A Shift in Perspective on Personality. *Journal of Personality, 86*, 69-82. <u>https://doi.org/10.1111/jopy.12319</u>
- McCormick, C. B., Dimmitt, C. A. R. E. Y., & Sullivan, F. R. (2012). Metacognition, Learning, and Instruction. In *Handbook of Psychology* (2nd ed., Vol. 7). Wiley. <u>https://doi.org/10.1002/9781118133880.hop207004</u>
- McKenna, M., Cervone, D., Roy, A., & Burkett, C. (2021). Personality Coherence in Acts and Texts: Searching for Coherence within and beyond Trait Categories. *European Journal of Personality, 36*, 347-370. https://doi.org/10.1177/08902070211022131
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, *18*, 5-11. https://doi.org/10.3102/0013189X018002005
- Mischel, W. (1968). Personality and Assessment. Wiley.
- Mischel, W. (1973). Toward a Cognitive Social Learning Reconceptualization of Personality. *Psychological Review, 80,* 252. <u>https://doi.org/10.1037/h0035002</u>
- Mischel, W., & Shoda, Y. (1995). A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. *Psychological Review*, *102*, 246-268. <u>https://doi.org/10.1037/0033-295X.102.2.246</u>
- Molenaar, P. C. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back into Scientific Psychology, This Time Forever. *Measurement, 2*, 201-218. https://doi.org/10.1207/s15366359mea0204 1
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, *5*, 159-168. <u>https://doi.org/10.1207/s15327574ijt0502_4</u>
- Nicholas, M. K., McGuire, B. E., & Asghari, A. (2015). A 2-Item Short Form of the Pain Self-Efficacy Questionnaire: Development and Psychometric Evaluation of PSEQ-2. *The Journal of Pain*, 16, 153-163. <u>https://doi.org/10.1016/j.jpain.2014.11.002</u>
- Pearman, A., Lustig, E., Hughes, M. L., & Hertzog, C. (2020). Initial Evidence for the Efficacy of an Everyday Memory and Metacognitive Intervention. *Innovation in Aging*, 4, 500-501. <u>https://doi.org/10.1093/geroni/igaa057.1617</u>
- Pervin, L. A. (1994). A Critical Analysis of Current Trait Theory. *Psychological Inquiry*, 5, 103-113. <u>https://doi.org/10.1207/s15327965pli0502_1</u>
- Petzold, N., Landinez, L., & Baaken, T. (2019). Disruptive Innovation from a Process

View: A Systematic Literature Review. *Creativity and Innovation Management, 28,* 157-174. <u>https://doi.org/10.1111/caim.12313</u>

- Phan, H. P. (2009) Relations between Goals, Self-Efficacy, Critical Thinking and Deep Processing Strategies: A Path Analysis. *Educational Psychology, 29*, 777-799. https://doi.org/10.1080/01443410903289423
- Pintrich, P. R., & de Groot, E. V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology*, *82*, 33-40. <u>https://doi.org/10.1037/0022-0663.82.1.33</u>
- Raykov, T., & Marcoulides, G. A. (2000). A Method for Comparing Completely Standardized Solutions in Multiple Groups. *Structural Equation Modeling*, *7*, 292-308. https://doi.org/10.1207/S15328007SEM0702_9
- Renner, K. H., Klee, S., & von Oertzen, T. (2020). Bringing Back the Person into Behavioural Personality Science Using Big Data. *European Journal of Personality*, 34, 670-686. <u>https://doi.org/10.1002/per.2303</u>
- Ritchie, L., & Williamon, A. (2011). Measuring Distinct Types of Musical Self-Efficacy. *Psychology of Music, 39*, 328-344. <u>https://doi.org/10.1177/0305735610374895</u>
- Ritchie, L., & Williamon, A. (2012). Self-Efficacy as a Predictor of Musical Performance Quality. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 334-340. <u>https://doi.org/10.1037/a0029619</u>
- Rogers, C. R. (1940). The Processes of Therapy. *Journal of Consulting Psychology, 4,* 161-164. <u>https://doi.org/10.1037/h0062536</u>
- Schellings, G., & Van Hout-Wolters, B. (2011). Measuring Strategy Use with Self-Report Instruments: Theoretical and Empirical Considerations. *Metacognition and Learning*, *6*, 83-90. <u>https://doi.org/10.1007/s11409-011-9081-9</u>
- Schunk, D. H., & Gunn, T. P. (1986). Self-Efficacy and Skill Development: Influence of Task Strategies and Attributions. *The Journal of Educational Research*, *79*, 238-244. https://doi.org/10.1080/00220671.1986.10885684
- Schunk, D. H., & Usher, E. L. (2011). Assessing Self-Efficacy for Self-Regulated Learning. In D. H. Schunk, & B. Zimmerman (Eds.), *Handbook of Self-Regulation of Learning* and Performance (pp. 282-297). Routledge. <u>https://doi.org/10.4324/9780203839010</u>
- Scott, W. D., Paup, S., & Kirchhoff, C. (2021). Clinical Application of Social Cognitive Theory: A Novel Personality Assessment Procedure and a Case Study of Personality Coherence. *European Journal of Personality*, *36*, 371-390. <u>https://doi.org/10.1177/08902070211028362</u>
- Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and Validation. *Psychological Reports*, 51, 663-671. <u>https://doi.org/10.2466/pr0.1982.51.2.663</u>
- Short, S., & Ross-Stewart, L. (2008). A Review of Self-Efficacy Based Interventions. In S. Mellalieu, & S. Hanton (Eds.), Advances in Applied Sport Psychology (pp. 231-290). Routledge.
- Sireci, S. G. (2007). On Validity Theory and Test Validation. *Educational Researcher, 36,* 477-481. <u>https://doi.org/10.3102/0013189X07311609</u>
- Sireci, S. G. (2020). De-"Constructing" Test Validation. Chinese/English Journal of Educational Measurement and Evaluation, 1, Article 3. https://www.ce-jeme.org/journal/vol1/iss1/3
- Stewart, D. (1981). The Application and Misapplication of Factor Analysis in Marketing Research. *Journal of Marketing Research, 18,* 207-214. https://doi.org/10.1177/002224378101800105

- Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*, 99-103. <u>https://doi.org/10.1207/S15327752JPA8001_18</u>
- Sullivan, H. S. (1953). The Interpersonal Theory of Psychiatry. Norton.
- Tanaka, J. S. (1993). Multifaceted Conceptions of Fit in Structural Equation Models. In K.A. Bollen, & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 10-39). Sage.
- Thorndike, E. L. (1904). An Introduction to the Theory of Mental and Social Measurements. Teachers College Press. <u>https://doi.org/10.1037/13283-000</u>
- Tsai, M. J., Wang, C. Y., & Hsu, P. F. (2019). Developing the Computer Programming Self-Efficacy Scale for Computer Literacy Education. *Journal of Educational Computing Research*, 56, 1345-1360. <u>https://doi.org/10.1177/0735633117746747</u>
- Uher, J. (2013). Personality Psychology: Lexical Approaches, Assessment Methods, and Trait Concepts Reveal Only Half of the Story—Why It Is Time for a Paradigm Shift. *Integrative Psychological and Behavioral Science*, *47*, 1-55. https://doi.org/10.1007/s12124-013-9230-6
- Van Gog, T., & Scheiter, K. (2010). Eye Tracking as a Tool to Study and Enhance Multimedia Learning. *Learning and instruction*, 20, 95-99. https://doi.org/10.1016/i.learninstruc.2009.02.009
- Vandergrift, L., Goh, C. C., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The Metacognitive Awareness Listening Questionnaire: Development and Validation. *Language Learning*, 56, 431-462. <u>https://doi.org/10.1111/j.1467-9922.2006.00373.x</u>
- Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and Learning: Conceptual and Methodological Considerations. *Metacognition and Learning*, 1, 3-14. <u>https://doi.org/10.1007/s11409-006-6893-0</u>
- Velicer, W. F., & Fava, J. L. (1998). Affects of Variable and Subject Sampling on Factor Pattern Recovery. *Psychological Methods, 3,* 231-251. https://doi.org/10.1037/1082-989X.3.2.231
- Wellman, H. (1985). *The Child's Theory of Mind: The Development of Conscious Cognition.* Academic Press.
- Wells, A., & Cartwright-Hatton, S. (2004). A Short Form of the Metacognitions Questionnaire: Properties of the MCQ-30. *Behaviour Research and Therapy*, 42, 385-396. <u>https://doi.org/10.1016/S0005-7967(03)00147-5</u>
- Wise, J. B. (2007). Testing a Theory That Explains How Self-Efficacy Beliefs Are Formed: Predicting Self-Efficacy Appraisals across Recreation Activities. *Journal of Social and Clinical Psychology*, 26, 841-848. <u>https://doi.org/10.1521/jscp.2007.26.7.841</u>
- Yu D., & Hang C. C. (2010). A Reflective Review of Disruptive Innovation Theory. International Journal of Management Reviews, 12, 435-452. https://doi.org/10.1111/j.1468-2370.2009.00272.x
- Zimmerman, B. J., & Schunk, D. H. (Eds.) (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives.* Routledge.
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. *American Educational Research Journal*, 29, 663-676. https://doi.org/10.3102/00028312029003663
- Zohar, A., & Barzilai, S. (2013). A Review of Research on Metacognition in Science Education: Current and Future Directions. *Studies in Science Education, 49,* 121-169. https://doi.org/10.1080/03057267.2013.847261

Supplemental Material

	Min/Max	Mean (SD)
SEP	22/63	50.1 (8.79)
SE Job	0/100	82.5 (17.7)
MTQ	8/56	41.9 (8.18)
Strengths	0/20	9.9 (4.13)
SE Strengths	0/100	75.5 (1.36)
Weaknesses	0/14	5.46 (3.36)
SE Weaknesses	5/80	35.3 (1.65)

Descriptive statistics for SEP, SE Job, MTQ, the number of strengths/weaknesses, and the mean SE for strengths/weaknesses

Correlations between self-efficacy scores and personal strength for the 30 KAPA sub-items $\!\!\!^*$

	SE		SE		SE
S1	.406***	S11	.330***	S21	.495***
S2	.447***	S12	.646***	S22	.300***
S3	.361***	S13	.556***	S23	.245***
S4	.346***	S14	.338***	S24	.520***
S5	.354***	S15	.561***	S25	.574***
S6	.548***	S16	.491***	S26	.565***
S7	.536***	S17	.496***	S27	.531***
S8	.445***	S18	.489***	S28	.573***
S9	.552***	S19	.483***	S29	.523***
S10	.591***	S20	.297***	S30	.451***

Note*: * *p* < .001

Correlations between self-efficacy scores and personal weakness for the 30 KAPA sub-items *

	SE		SE		SE
W1	.619***	W11	.225***	W21	.195**
W2	.400***	W12	.224***	W22	.319***
W3	.145**	W13	.504***	W23	.205**
W4	.260***	W14	.403***	W24	.226***
W5	.417***	W15	.483***	W25	.525***
W6	.267***	W16	.394***	W26	.539***
W7	.351***	W17	.350***	W27	.369***
W8	.382***	W18	.232***	W28	.441***
W9	.474***	W19	.372***	W29	.343***
W10	.449***	W20	.171**	W30	.377***

Note* ** *p* < .01, * *p* < .001.



Figure representing situations from KAPA where personal strength/weakness was rated most relevant for three teachers who all had the same self-efficacy score and share the same strength of being determined.

Self-efficacy for Performing Questionnaire (SEP)

We would like you to consider the main task of your job/profession/specialism. What is your job/profession/specialism (e.g. lifeguard, teacher, carer, engineer)?

Please name a main task for your job/profession/specialism: _

We would like you to consider the next questions with that main task in mind. How confident are you that you can successfully carry out that task?

Can you describe why you have chosen this level of confidence for this particular task in words?

Now, please indicate how much you agree or disagree with each of the following statements, specially regarding your confidence in how you will perform during this activity.

	Not at all Sure 0%	-	-	-	-	-	Completely 100%
1. I am confident that I can successfully carry out the task.	1	2	3	4	5	6	7
2. I have set important goals to attain during this task, but I cannot achieve them.	1	2	3	4	5	6	7
3. I am likely to avoid (not face) difficulties and challenges during the task itself.	1	2	3	4	5	6	7
4. If I perceive the events or context surrounding this task to be too stressful, I cannot even attempt it.	1	2	3	4	5	6	7
5. If something unexpected happens during the task, I can handle it well.	1	2	3	4	5	6	7
6. I am likely to avoid this task if it appears too difficult for me.	1	2	3	4	5	6	7
7. I feel insecure about my skills for this task.	1	2	3	4	5	6	7
8. I am likely to give up easily during the task.	1	2	3	4	5	6	7
9. I am capable of dealing with problems that might come up during the task.	1	2	3	4	5	6	7

	Very untrue of me	Mostly untrue of me	Sometimes f untrue of me	Neither true or untrue	Sometimes true of me	Mostly true of me	Very true of me
1. I think about and consider the before and after context of the task I am engaged in.	1	2	3	4	5	6	7
2. I actively consider elements of my strategic engagement with tasks.	1	2	3	4	5	6	7
3. I reflect on events in relation to my past experiences and future goals.	1	2	3	4	5	6	7
4. I actively think about how I do things.	1	2	3	4	5	6	7
5. I notice my thinking when I decide on or do something.	1	2	3	4	5	6	7
6. I plan my thinking while doing things, practising reflection in action.	1	2	3	4	5	6	7
7. I evaluate and organise my thinking in relation to what I have done.	1	2	3	4	5	6	7
8. I examine my thoughts during tasks.	1	2	3	4	5	6	7

30 Contextual Items (in this iteration) of the Knowledge and Appraisal method of Personality Architecture (KAPA)

1. Worry and become anxious before going to a social event where you don't know too many people.

2. Become angry with someone at a store if they don't do a good job helping you.

3. If you get some bad news, figure out a way to avoid becoming discouraged and depressed.

4. If you make a mistake, manage to avoid feeling inferior to other people.

5. Even if you're trying to avoid eating too much, give in to impulses and overeat.

6. Remain calm and "cool-headed" during a time of stress.

7. If you are meeting a group of strangers, smile and be socially outgoing.

8. If someone you know throws a big, loud party, avoid going.

9. If you are in a work/discussion group, sit back and let the others do the talking, rather than speaking up yourself.

10. If you have a free weekend coming up, plan a series of "high-energy" events.

11. If you are at an amusement park/theme park, skip riding the biggest fastest rides.

12. Literally jump for joy when hearing very good news.

13. Spend a lot of time during the day just daydreaming.

14. Become bored if you are asked to sit through a long presentation/demonstration.

15. Spend a lot of time during the day analysing your own feelings.

16. If some friends are thinking of getting or cooking an unusual foreign "exotic" food, for dinner, suggest something simpler and more well-known to eat instead.

17. Become bored if you are listening to people having a long, complex argument.

18. Remain tolerant of other people's lifestyles, even if you think their lifestyle is not a good way to live.

19. Become suspicious if someone you have just met offers to do something nice for you.

20. If you need something from a friend, trick them into giving it to you.

21. If someone tells you about a charity that sounds like a good cause, make a generous donation to that charity.

22. If you spend a long weekend with your relatives or people close to you, avoid getting into any arguments with them.

23. Avoid bragging about your own talents and accomplishments when talking to friends.

24. If someone on the street asks you for money, respond to them in a sympathetic manner.

25. Keep yourself fully prepared for your work week by organising and doing things in advance.

26. Keep your personal belongings at home neat and well-organised.

27. If you are feeling a little tired and would like the day off from work, call in sick (even if you are not literally sick).

28. Stick to a self-improvement programme for a long period of time.

29. If you have to produce a report for your work, waste a lot of time before you get around to actually starting on the writing.

30. Before making an important decision, think about all the options and the consequences of different options before deciding.