ANZSOG

Psychometric testing for recruitment and selection in the NSW Public Service

ANZSOG

PSYCHOMETRIC TESTING FOR RECRUITMENT AND SELECTION IN THE NSW PUBLIC SERVICE

A thematic literature review for the Office of the Public Service Commissioner

12 November 2025
Dr Andrew Bushnell,
Dr Natalia Rajendran,
Dr Patrick Brownlee,
Dr Lisa Carson

ΛNZSOG

CONTENTS

ANZSOG

TABLE OF TABLES

EXECUTIVE SUMMARY

The Office of the Public Service Commissioner (OPSC) of New South Wales (NSW) engaged the Australia and New Zealand School of Government (ANZSOG) to undertake a **thematic literature review** of psychometric testing used in merit selection processes. A focus area for the review was on predictive validity and bias considerations in such tests. Application to the public sector was also investigated.

The review is based on an iterative key word search and targeted searches of academic databases, journals, grey literature and government websites. It is informed by earlier work commissioned by OPSC in 2017, with the intention of providing an updated understanding of empirical evidence.

Main findings

The **main findings** of the review are organised by the 4 types of psychometric test used in the NSW public sector. The findings are presented below as **practice points** distilled from an analysis of the literature

Cognitive ability

- Recent research reaffirms the moderate predictive validity of cognitive ability testing for job performance; however, the correlation is lower than previously understood.

- General cognitive ability tests should be part of a larger process that uses additional selection methods.

- Testing specific mental abilities (and those relevant to jobs) can add predictive validity.

- Recent evidence finds possible effects of linguistic and cultural proficiency on the predictive validity of (general) cognitive ability testing.

- Predictive validity varies across occupations; selection based upon test scores should not aim to seek comparisons across objectively distinct job categories to establish practice norms (e.g., for test administration guidance or thresholds).

Personality

- Of the Big Five personality traits, conscientiousness is a moderate predictor of job performance. The other 4 traits are weak predictors. The Big Five do not have incremental validity with each other, meaning testing combination of these constructs does not provide added or cumulative value. This suggests that testing for conscientiousness is more worthwhile than a full personality inventory.

- By contrast, there is evidence that some personality traits, including conscientiousness, have incremental validity over cognitive ability. So, it is worthwhile to test for both. In general, personality testing is best used in conjunction with other test types.

- Job characteristics moderate the predictiveness of personality traits for job performance in various ways, including the finding that personality traits are more important for less structured roles. But conscientiousness is the only trait found to predict performance widely across job roles (and then only moderately).

- Personality testing for traits (like the Big Five) is more predictive for job performance than testing based on types (like the popular Myers-Briggs Type Indicator). Predictive validity can also be increased by contextualising the test within the occupational requirements.

ΛΝΖSOG

- Culture has also been found to influence personality testing: while the underlying traits are generally consistent across people from different cultures, how they express those traits is influenced by culture and this bears on how tests are designed and ought to be interpreted.

- Integrity tests measure personality traits associated with counterproductive behaviour and range from moderate-to-strong predictors of job performance and shown to be especially useful for selection in law enforcement and customer service roles.

Emotional intelligence (EI)

- Recent evidence supports the moderate predictive validity of both ability and trait EI for various work-related outcomes, including job performance. Ability and trait EI based on self-reports are more predictive than ability EI testing based on performance.

- There is some evidence that EI is predictive of at least some work-related outcomes over and above cognitive ability and personality testing. EI is most predictive of managers' organisational commitment and may also increase their job performance. Importantly, managers' EI has also been found to moderate the effect of employees' EI on their performance. Manager EI supports overall effectiveness.

- The literature supports the value of EI in public sector contexts, finding that it affects public sector motivation and work-related outcomes for public servants.

- As with personality tests, EI testing across cultures has tended to find that while the underlying abilities and traits are consistent across cultures, their expression varies, meaning that caution is required in the interpretation of their results.

Situational Judgement Tests (SJTs)

- Recent research supports the use of SJTs in selection processes. While SJTs are highly heterogeneous, being developed for specific situations, they tend to be reliable and demonstrate moderate predictive validity for job performance.

- Knowledge-based tests (that ask candidates what they should do in a given scenario) tend to correlate with cognitive ability, whereas behaviour-based tests (that ask candidates what they would do) tend to correlate with personality traits. But SJTs have been shown to have incremental validity over and above these other tests.

- SJTs have been found to be useful for screening candidates, as they are about as predictive of job performance as structured interviews but discriminate better between average and low performers than between high performers.

- SJTs have shown differential results across groups, tending to favour females and members of cultural majorities. Culture is a potential confounder of results and may be relevant to how tests are designed and conducted.

Key themes

Five **key cross-cutting themes** emerge from the findings:

1. Psychometric tests are most useful as part of a larger, multi-stage selection process.

2. Psychometric testing shapes the candidate pool and so the potential effects of inherent biases and differential results across groups requires careful monitoring.

3. The use case for non-cognitive psychometric tests is strongest for leadership and managerial roles.

4. The development of artificial intelligence is changing the way that selection operates in practice.

5. Recent literature supports the usefulness of psychometric testing for recruitment and selection – and largely support current NSW policy.

Headline developments and future work

In conclusion, there are 3 **headline developments** since OPSC last reviewed the evidence base in 2018:

1. Cognitive ability testing remains the strongest psychometric tool for predicting job performance, but new evidence finds that this relationship is weaker than previously understood and supports the incremental validity of narrow cognitive abilities in some circumstances.

2. Regarding non-cognitive psychometric testing, recent evidence confirms that conscientiousness is the main personality-trait predictor of job performance, while strengthening the importance of this trait and emotional intelligence for management roles and confirming the reliability and moderate predictiveness of SJTs.

3. The academic literature identifies a range of possible biases and confounds affecting psychometric testing, especially how test results are interpreted and incorporated in selection processes, and emergent challenges around faking. Culture is usually demarcated between collectivist and individualist societies, or between black minority and white majority. Australia's context has unique features.

Regarding **future research**, there remains a lack of public sector specific research, especially for Australia. Given the need to match selection processes to specific job descriptions and noting unique aspects of public sector employment (e.g. integrity requirements, public sector motivation, and hierarchical organisation structures), ANZSOG identifies an opportunity for governments undertake further research on the direct use and utility of psychometric using internal HR and performance data.

To assist NSW OPSC in practice, Appendix A maps the findings from this ANZSOG report onto the NSW Public Sector Capability Framework (2020), on which job descriptions are based, providing guidance for the composition of selection processes, and reinforcing the usefulness of these tests.

INTRODUCTION

The NSW Office of the Public Service Commissioner (OPSC) engaged ANZSOG to undertake a thematic literature review on the use of psychometric testing in recruitment and selection. This review also examines the predictive validity of different tests, their applicability across roles, their role in broader selection processes, and potential biases. The report provides an updated analysis of key findings from academic literature, building on earlier work (Griffin, 2018).

Psychometric tests assess constructs like intelligence, personality, and competencies, offering an objective comparison of candidates. In recruitment and selection, psychometric testing is widely used to enhance objectivity and consistency (Carless, 2009). The human resource (HR) field generally supports its use, as it helps organisations meet legal obligations and is cost-effective (Potočnk et al, 2021). However, research also highlights challenges, including validity concerns and potential biases related to gender, ethnicity, culture, language, and disability (Rust et al, 2009; Bratton et al, 2022). Research has contested whether test results reflected stable traits or situational manifestations of those traits (Rust et al, 2009).

NSW government agencies use various psychometric tests. As of 2021, 84 tests across 15 categories were approved by the NSW Government (according to supplied information). Selection processes are governed by the *Government Sector Employment Act 2013*, which mandates a merit-based, apolitical public service (s 6). Related regulations require at least two capability assessments, including an interview (*Government Sector Employment (General) Rules 2014*, ss 16-18). The *NSW Government Recruitment and Selection Guide* identifies 4 psychometric test types: cognitive ability, personality, emotional intelligence, and situational judgment. It also prohibits using test results as the sole basis for hiring decisions (Public Service Commission, 2025).

METHOD

The findings below are based on a rapid review of recent academic and grey literature over a two-month period in 2025. In collaboration with NSW OPSC, a set of search terms was developed based on a broad scan of the field, an earlier paper written on this topic for OPSC (Griffin, 2018) and the agreed terms of the review.

Table 1: Search terms – Theme 1: Psychometrics

Aspect	Search terms
Objects of study	Psychometric test*, cognitive ability, general intelligence, personality, emotional intelligence, situational judgement
Testing	Assess*, interview, exam, situation
Validity	Valid*, predict*, perform*, sound*, outcome, confidence

Table 2: Search terms – Theme 2: Practical challenges

Challenge	Search terms
Confounds	Bias, rac*, ethnic, indigenous, First Nations, Aboriginal, gender, sex, socioeconomic, class, education level, disability
Processes	Selection, recruit*, promotion, stage
Applications	Public service, public sector, civil service, government, position, job role, management, leadership

These terms were used in an iterative search process of academic literature, with a primary interest in meta-analyses. First, via the University of Melbourne's EBSCO search tool and Google Scholar published from 2015 up to and including 2025. Secondly, targeted searches were conducted: selected journals and government sites, 'forward citation mapping' of meta-analyses cited in the earlier Griffin (2018) report, and by using the tools ChatGPT and Perplexity to address gaps and to search inside articles. The background sections were compiled from targeted searches of handbooks and reviews.

Captured articles were sorted by relevance into the different types of tests and by theme. Empirical studies bearing on the question of predictive validity were separated from literature reviews and analyses. ChatGPT and Perplexity were used in the editing of the body of the report (excluding practice and discussion points) to suggest revisions for conciseness that were then checked and further revised by the authors.

As with all methods, there are limitations. For this study, the rapid reporting requirement and, as the research revealed, a paucity of studies that identify the public service as a unique context. Studies in Australia are also not immediately apparent which is significant given the emphasis in testing and job context on culture and on societal meta-concepts (such as collectivist and individualist). Much of the literature focuses on refining known measures and building the evidence base. No evidence was available from psychometric test providers about their validity claims. Finally, note that predictive validity represents a set of correlations rather than causal relationships between the constructs and performance outcomes.

FINDINGS

This section provides a thematic review of evidence and issues related to each of the 4 types of psychometric tests endorsed by NSW Government policy. To aid practitioners, each sub-section begins with practice points distilled from the research findings.

1. Cognitive ability

Practice points

- Recent research reaffirms a moderate predictive validity of cognitive ability testing for job performance; however, the correlation is lower than previously understood.

- General cognitive ability tests should be part of a larger process that uses additional selection methods.

- Testing specific mental abilities (and those relevant to jobs) can add predictive validity.

- Recent evidence finds possible effects of linguistic and cultural proficiency on the predictive validity of (general) cognitive ability testing.

- Predictive validity for cognitive ability testing varies across occupations; selection based upon test scores should not aim to seek comparisons across objectively distinct job categories to establish practice norms (e.g., for test administration guidance or thresholds).

Background

General cognitive ability (GCA), sometimes called general intelligence or general mental ability, is broadly understood as the ability to learn and apply new information through a process of abstraction (Salgado, 2017). Cognitive ability testing assesses candidates' general intelligence ('g') and related abilities like verbal and numerical reasoning (Beaujean & Benson, 2019; Deary, 2013). The construct g is generally accepted as picking out a real trait or personal attribute (Gottfredsson, 1994; Deary, 2012). Cognitive ability test results have for a long time been considered highly predictive of job performance and are widely used in recruitment and selection processes (Schmidt & Hunter, 1998; Salgado, 2017).

Predictive validity

Recent meta-analyses **confirm a moderate predictive validity** for GCA tests – but the relationship is **weaker than previously understood**. Sackett et al. (2024) analysed 153 studies (2000–2021) correlating GCA with overall job performance ($\rho = 0.22$) and task performance ($\rho = 0.23$). A prior study (Sackett et al., 2022) found that a refined methodology applied to older datasets reduced GCA validity from 0.51 to 0.31. Notably, objective performance measures ($\rho = 0.38$) correlated more strongly than supervisor ratings ($\rho = 0.21$).

Salgado & Moscoso (2019) similarly critiqued past methodologies in their meta-analysis ($N = 630$) of the General Aptitude Test Battery (12 cognitive tests) for medium-complexity jobs (62% of U.S. employment). GCA predicted job performance across supervisor ratings, production records, work sample tests, and training outcomes but at varying strengths. **It was more predictive for complex jobs and training success than job performance**, aligning with Sackett et al. (2024). While GCA remains a valid predictor, **the authors advise using it alongside other selection methods for greater practical relevance**.

On this last point, recent research suggests that **assessing narrow cognitive abilities** and **tailoring tests to job-specific mental skills** can enhance predictive validity. Nye et al. (2022), a meta-analysis of studies from 1990–2020, found that while GCA correlates with multiple job performance dimensions, the strength of these relationships varies. Findings demonstrate that narrow cognitive abilities can add incremental validity to GCA testing. The largest gains come from abilities weakly correlated with GCA, such as processing speed, auditory and visual processing, and reaction and decision speed. This suggests that cognitive testing should incorporate job-relevant, lower *g*-loaded abilities where applicable. Grobelny (2018) also found that weighting specific mental abilities based on job requirements improves predictive accuracy, with occupation acting as a moderator of GCA's validity. These findings support **aligning assessments with the specific cognitive demands of different roles**.

Bias and confounds

Cognitive ability tests raise concerns about bias, particularly the risk that cultural assumptions may affect their predictive validity across different groups. While Australian evidence is scant, international research highlights potential issues. In the US, evidence suggests that cognitive tests may overpredict job performance for African Americans (Berry & Zhao, 2015). Conversely, a UK meta-analysis (Te Nijenhuis, Pesta & Fuerst, 2024) found that white candidates outperformed non-whites on GCA tests, though this gap was smaller for younger cohorts born in the UK, suggesting possible influences of language proficiency and acculturation. **Despite concluding that GCA testing is "valid and unbiased," they acknowledge concerns about adverse effects on some groups.** Combined with Sackett et al.'s (2022) findings, for the authors this suggests that GCA testing is best used within a broader selection process that carefully integrates multiple assessment methods. More research is needed to identify generalised cultural bias effects.

2. Personality

Practice points

- Of the Big Five personality traits, conscientiousness is a moderate predictor of job performance. The other 4 traits are weak predictors. The Big Five do not have incremental validity with each other, meaning testing combination of these constructs does not provide added or cumulative value. This suggests that testing for conscientiousness is more worthwhile than a full personality inventory.

- By contrast, there is evidence that some personality traits, including conscientiousness, have incremental validity over cognitive ability. So, it is worthwhile to test for both. In general, personality testing is best used in conjunction with other test types.

- Job characteristics moderate the predictiveness of personality traits for job performance in various ways, including the finding that personality traits are more important for less structured roles. But conscientiousness is the only trait found to predict performance widely across job roles (and then only moderately).

- Personality testing for traits (like the Big Five) is more predictive for job performance than testing based on types (like the popular Myers-Briggs Type Indicator). Predictive validity can also be increased by contextualising the test within the occupational requirements.

- Culture has also been found to influence personality testing: while the underlying traits are generally consistent across people from different cultures, how they express those traits is influenced by culture and this bears on how tests are designed and ought to be interpreted.

- Integrity tests measure personality traits associated with counterproductive behaviour and are moderate-to-strong predictors of job performance and shown to be especially useful for selection in law enforcement and customer service roles.

Background

Personality in psychometrics refers to stable traits that shape thoughts, feelings, and behaviour (Hughes & Batey, 2017). The Five Factor Model ("Big Five")—openness, conscientiousness, extraversion, agreeableness, and neuroticism—is widely accepted, though there are alternatives like HEXACO, which has 6 factors. These traits, while stable in adulthood, can shift with age and life events (Roberts, Walton & Viechtbauer, 2006). In recruitment, personality traits have been correlated with job satisfaction, performance, leadership, and counterproductive behaviour. Their predictive power improves when assessments target job-relevant traits. However, like other psychometric tools, personality tests are typically used within a broader selection process (Leyens, Ritz & Hofmans, 2024).

Predictive validity

Recent research finds that **conscientiousness is a moderate predictor of job performance**. But the other 4 of the Big Five traits correlate only weakly with job performance.

Zell & Lesick (2022) conducted a second-order meta-analysis of 54 studies, confirming significant but weak-to-moderate correlations. Conscientiousness is the strongest job performance predictor among the Big Five, but only shows a barely moderate relationship ($p = 0.20$). Additionally, as Table 2 shows, predictiveness for all traits varies across performance domains, with the relationships differing between job and academic performance.

Table 3: Big 5 personality traits vs performance (extracted from Zell & Lesick, 2022)

Domain	Extraversion	Agreeableness	Conscientious-ness	Neuroticism	Openness
Overall	0.10	0.10	0.19	-0.12	0.13
Job (m=32)	0.14	0.11	0.20	-0.15	0.11
Academic (m=7)	-0.01	0.07	0.28	-0.03	0.14
Other (m=15)	0.08	0.09	0.12	-0.10	0.16

Another study of 4 different proposed interactions *between* the Big Five personality traits found no significant incremental predictive validity Bradburn et al. (2020). This finding suggests that these traits should be assessed as required based on job description, because there is no significant additive effect of using a full inventory.

Conversely, some personality traits have incremental validity for job performance over and above cognitive ability. Iliescu et al. (2023) found significant incremental validity for conscientiousness, neuroticism, and agreeableness. For selection, these findings support testing these traits in addition to cognitive ability.

Research shows **job context moderates the personality-performance link** in various ways (noting that this link ranges from weak-to-moderate anyway):

- A meta-analysis of Big Five inventories specifically using a forced-choice format – found various relationships between personality traits and performance in different occupations (Salgado, Anderson & Tauriz, 2015). Conscientiousness was found to be a valid predictor for the 5 occupational groups tested (customer service, managerial, police, sales, and supervisor) with a highest value for sales (.27). Other traits were only weak predictors at best – and openness was not found to be a valid predictor at all.

- Judge & Zapata (2015) found that all Big Five traits predict performance better in unstructured jobs requiring more agency, and that job context factors "activate" traits, making them more predictive.

- He, Donnellan & Mendoza (2019) showed openness and agreeableness predict customer service performance and found stronger than average correlations between extraversion and openness and manager job performance. But they found little support for broader job-specific correlations.

- Bergner (2020) confirmed extraversion predicts leadership success (more strongly than cognitive ability).

Integrity tests

Integrity tests assess traits like honesty, reliability, and ethics, predicting workplace behaviour. There are two types: covert (or personality-based) tests, which assess personality traits indirectly and add validity beyond the Big Five, and overt tests, which directly ask about past behaviour and strongly correlate with HEXACO's Honesty-Humility dimension (He & Gong 2010; Casillas et al, 2009; Marcus, Lee & Ashton, 2007).

These tests predict counterproductive work behaviour (CWB) or workplace deviance (which includes rule-breaking intended to benefit the organisation) with moderate-to-strong effects. A meta-analysis on workplace deviance (Wing-Man Lau et al., 2023) found an overall validity of .29, rising to .43 when corrected, with the strongest effects in military, law enforcement (ρ = .50), and customer service roles (ρ = .47). The effect sizes across countries varied, but there were no significant differences between sub-groups within nations for which data was available.

Integrity tests also correlate with job performance (Griffin, 2018).

When used in combination with other tests, general cognitive ability has only a weak correlation with integrity test scores (Cunningham et al, 2018).

Bias and confounds

Personality testing raises several key considerations for selection and recruitment, four of which are outlined here. First, a longstanding finding is that trait-based models like the Big Five are more predictive than type-based models such as the Myers-Briggs Type Indicator (MBTI) (McCrae & Costa, Jr., 1989; Pittenger, 1993). A more recent review (Randall, Isaacson & Ciro, 2017) found evidence supporting MBTI's construct validity and test-retest reliability, but did not compare MBTI with other instruments directly nor challenge the earlier consensus.

Second, **contextualised personality assessments** that tailor questions to job roles **improve predictive validity**. A study on law enforcement recruitment (Fisher et al, 2017) found that a contextualised conscientiousness measure (r = .30) was far more predictive of job performance than a general version (r = -.01). That finding supports earlier meta-analysis showing contextualised personality scores nearly double the validity of generalised test scores for four of the Big Five traits (cited in Sosnowska et al, 2018).

Third, test-taker faking is a concern. There is some evidence that **faking sometimes works** (Kasten, Freund & Staufenbiel, 2020). Instrument design can mitigate this (Lee, 2023). Situational Judgment Tests (SJTs) (see below) also resist faking better than personality inventories because faking an SJT correlates with cognitive ability, while faking a personality test aligns with personality traits (Kasten, Freund & Staufenbiel, 2020).

Finally, **cultural differences affect personality test results**. While personality traits are comparable across cultures, cultural norms shape the value placed on different traits, influencing test responses and interpretation. A second-order meta-analysis (He, Donnellan & Mendoza, 2019) found conscientiousness and emotional stability showed little cross-cultural variation, but agreeableness, openness, and extraversion differed between collectivist (China, Taiwan, Singapore) and individualist (USA, Europe) cultures. These differences could affect the weighting of personality facets, particularly for global firms. A Turkish meta-analysis (Askun et al., 2024) found stronger

effects of agreeableness and openness in collectivist settings and a weaker effect for emotional stability when compared to studies in individualist countries. More narrowly, cross-cultural differences also impact faking. Lee's (2023) study comparing US and South Korean test-takers found no significant difference in overall faking tendencies, though cultural norms influenced which traits were faked. These results suggest a need for care in the interpretation of personality testing in multicultural workforces.

3. Emotional intelligence (EI)

Practice points

- Recent evidence supports the moderate predictive validity of both ability and trait EI for various work-related outcomes, including job performance. Ability and trait EI based on self-reports are more predictive than ability EI testing based on performance.

- There is some evidence that EI is predictive of at least some work-related outcomes over and above cognitive ability and personality testing. EI is most predictive of managers' organisational commitment and may also increase their job performance. Importantly, managers' EI has also been found to moderate the effect of employees' EI on their performance. Manager EI supports overall effectiveness.

- The literature supports the value of EI in public sector contexts, finding that it affects public sector motivation and work-related outcomes for public servants.

- As with personality tests, EI testing across cultures has tended to find that while the underlying abilities and traits are consistent across cultures, their expression varies, meaning that caution is required in the interpretation of their results.

Background

Emotional intelligence (EI) gained prominence in the 1990s and is widely used in selection. Broadly, EI can be a test of ability focusing on skills like perceiving and managing emotions, or traits like those tested in personality testing, or both. Scholars generally favour ability EI conceptually as mixed models rely on self-reports, diverging from traditional intelligence measures (Rivers et al, 2020; Robinson, 2024; Lievens & Chan, 2017). Ability EI has a weak positive correlation with workplace performance, over and above cognitive ability and personality testing. Proposed mechanisms for the effect include teamwork, stress tolerance, and leadership. Some EI abilities may be malleable in adults, but debates persist about its relationship with general intelligence, personality, brain functions, and cross-cultural validity (Robinson, 2024).

Predictive validity

Two recent meta-analyses support the predictive validity of EI testing (of all types) for various work-related outcomes.

Doğru (2022) undertakes a meta-analysis of studies conducted between 1990-2020 linking EI with organisational commitment (acceptance of rules and values), organisational citizenship behaviour (meeting and exceeding duties), job performance, job satisfaction, and a relationship to (reduced) job stress. Overall, the meta-analysis finds moderate correlations between all 3 kinds of EI (ability, self-report, and mixed) and all 5 workplace outcomes. The strengths vary, as shown in Table 3, below, with the strongest effects being on organisational citizenship and job stress.

Table 4: Emotional intelligence vs workplace outcomes (extracted from Doğru, 2022)

Type	Organisational commitment	Organisational citizenship	Job performance	Job satisfaction	Job stress
Overall	0.26	0.36	0.30	0.29	-0.43
Ability	0.22	0.29	0.28	0.24	-0.42
Self-report	0.28	0.37	0.33	0.31	-0.45
Mixed	0.27	0.35	0.31	0.30	-0.37

Similarly, Grobelny, Radke & Paniotova-Macza (2021) analysed 116 studies on EI and found robust predictive validity for various EI test types and job performance (ρ = 0.33). The correlation ranged from 0.31 for performance-based ability tests, to 0.38 for self-reported trait tests, to 0.48 for self-reported ability tests – these latter instruments show strong predictive validity, particularly for abilities over traits. Combined, the two studies suggest a robust relationship between EI and various work-related outcomes, testable by different instruments.

Regarding incremental validity for job performance, Grobelny, Radke & Paniotova-Macza (2021) find previous studies on this to be flawed and cannot be assessed based on their datasets. But there is some evidence regarding incremental validity for other work-related outcomes. A meta-analysis by Miao, Humphrey & Qian (2017) found that self-report and mixed EI tests show incremental validity for **organisational commitment** over and above cognitive ability and the Big Five, with self-report EI also adding validity for **turnover intentions**. All types of EI testing are also correlated with job satisfaction. To the extent that these outcomes relate to job performance this study provides some indirect evidence for the usefulness of EI testing for that purpose.

Moderating effects of occupations and industries

Both Doğru (2022) and Grobelny, Radke & Paniotova-Macza (2021) highlight ways that **managerial roles moderate EI's predictive validity**. Doğru (2022) finds that EI is more predictive for managers in organisational commitment and job performance but less so for other work outcomes. For non-managers, EI has stronger effects on organisational citizenship behaviour, job satisfaction, and job stress. Grobelny, Radke & Paniotova-Macza (2021) report a slightly weaker EI-job performance correlation for managers (ρ = .30) compared to their overall finding (.33). These results suggest recruiters should weigh EI differently based on whether a role involves management.

Grobelny, Radke & Paniotova-Macza (2021) also examine EI's predictive value by industry. In the context of their overall finding of good predictive validity, they also show that EI is most predictive in call centres (.60 – a strong predictor here), financial services (.44), uniformed services (.42), and manufacturing (.44). Its impact is weaker in healthcare (.16), telecommunications (.16), and retail .18). **Government and public services tends towards good (.33)** and ranks in the mid-range of these findings

On this last point, other recent research indicates EI's importance in public sector roles. Levitats et al. (2019) found that high EI among managers enhances employee engagement, while EI training improves engagement and social responsibility. Triyono & Tjahjono (2023) show that public service motivation mediates the positive relationship between EI and organisational commitment. Studies of public servants from Ethiopia (Wonda, 2024) and the Netherlands (Eshuis et al., 2023) further link EI dimensions – such as social awareness and emotion appraisal – to improved job performance and other work-related outcomes like efficiency, teamwork, and stress management.

Overall, EI influences work outcomes across industries and hierarchies, with stronger effects in certain roles and sectors. This highlights the need for tailored approaches to EI assessment during recruitment and development programs.

Bias and confounds

As with personality tests, the evidence for the cross-cultural validity of EI tests is mixed. A study of Brazilian and Portuguese university students (Pereira Teques et al, 2023) finds that an EI instrument demonstrated similar psychometric properties (construct validity, reliability, and criterion validity) for both countries. However, a larger, older study (Karim & Weisz, 2010) comparing French (Western, individualist) participants with Pakistani (Eastern, collectivist) participants found significant group differences in scores. The study found that the test successfully measures the same construct across cultures, and that this construct is distinct from cognitive ability and personality – though not predictive over and above them. However, individualist cultural background correlates with higher scores in all branches of the test (perceiving, using, understanding, and managing emotions). The authors **caution applying the test across cultures, especially where the scoring norms are non-local** (in this case, based on American expert consensus).

Genos Emotional Intelligence Inventory (used in NSW Public Service)

The evidence is **broadly supportive of the reliability and predictive validity of these tests used in NSW government**. The Genos Emotional Intelligence Inventory (GEII), developed in Australia, is a 7-factor measure designed for workplace applications. A 2010 journal article by Genos's Director Research (Gignac, 2010) shows that the test is reliable. However, the author notes that *reliability* is higher for the total score than for the subscale scores, implying that the former is suitable for use in selection and recruitment processes (along with predictive validity data) and the latter scores are only suitable for learning and development contexts. Previous peer-reviewed studies also correlate GEII with a range of workplace outcomes for pharmaceutical sales representatives' job performance (revenue, days out on calls, short and long sales calls) (Palmer et al, 2009). For the EQi 2.0, a recent South African study (van Lill et al, 2023) finds that the model does reliably measure a single construct of EI (contrary to some earlier findings – van Zyl, 2014) and that this construct is moderately correlated with job performance.

4. Situational Judgement Tests (SJTs)

Practice points

- Recent research supports the use of SJTs in selection processes. While SJTs are highly heterogeneous, being developed for specific situations, they tend to be reliable and demonstrate moderate predictive validity for job performance.

- Knowledge-based tests (that ask candidates what they should do in a given scenario) tend to correlate with cognitive ability, whereas behaviour-based tests (that ask candidates what they would do) tend to correlate with personality traits. But SJTs have been shown to have incremental validity over and above these other tests.

- SJTs have been found to be useful for screening candidates, as they are about as predictive of job performance as structured interviews but discriminate better between average and low performers than between high performers.

- SJTs have shown differential results across groups, tending to favour females and members of cultural majorities. Culture is a potential confounder of results and may be relevant to how tests are designed and conducted.

Background

Situational Judgment Tests (SJTs) evaluate judgment through job-related scenarios, correlating with job performance, leadership, and interpersonal skills (Whetzel et al, 2020). Their predictive validity depends on factors like response format, scenario design, and rating methods (Whetzel et al, 2020). High-fidelity SJTs, which simulate job scenarios, better predict performance but are less practical than low-fidelity questionnaire formats (Corstjens, Lievens & Krumm, 2017). A key debate is whether SJTs assess job-specific competencies or general traits like judgment. Contextualised SJTs outperform general measures, though single-factor SJTs also predict performance effectively (Corstjens, Lievens, Krumm, 2017; Chamorro-Premuzic & Furnham, 2010). Group differences in SJT outcomes arise from cultural influences, cognitive abilities, and personality traits, highlighting the need to consider these factors in test design and interpretation (Lievens, 2005; Whetzel et al, 2020).

Predictive validity

Recent research supports a moderate predictive validity of SJTs for job performance. Webster et al. (2020) conducted a meta-analysis of 30 studies on SJTs in medical school admissions, finding a moderate-to-good predictive validity for job performance (ρ = .32) and incremental validity beyond cognitive ability and personality. Despite sample heterogeneity, they conclude SJTs are valuable early in selection, given their low cost, ease of use, and ability to differentiate average from low performers - though they struggle to distinguish between top candidates. McDaniel et al. (2007) originally established that SJTs were moderately predictive of job performance (ρ = .26) and offering **incremental validity beyond cognitive ability and personality**.

SJTs are highly heterogeneous, often designed for specific jobs and tracking different constructs, with flexibility considered a strength. While they show low internal consistency, SJTs demonstrate strong test-retest reliability, justifying their continued use (Kepes et al, 2024; Kasten & Freund, 2020; Harenbock, Forthmann & Holling, 2023). Consistent with this, Christian et al. (2010) found that **SJTs' predictive validity improves when constructs align with job-specific performance facets**. Tiffin et al. (2019) speculate that construct-driven SJTs that measure generic constructs like integrity might be more suitable for early-process screening, whereas more job-specific SJTs might be suited to later-stage training after exposure to work scenarios.

Bias and confounds

SJTs show differential results across groups, raising concerns about potential biases in their design. Bardach et al. (2021) examined SJTs in UK teacher education selection, finding that females consistently outscore males, with the gap larger among ethnic minorities. White candidates outperform non-whites across all domains, with the greatest disparity for non-white males; socioeconomic status had no significant effect. Whether these patterns hold in other job contexts remains unclear. In the US, Rosales, Conley & Norris (2024) also found an SJT in an anaesthesiology residency program favoured white applicants over Black and Hispanic applicants and those from US medical schools. In the UK, Lievens et al. (2016) found SJTs in medical and dentistry admissions also showed differences by socioeconomic status, sex, and ethnicity and recommend weighting SJTs equally with cognitive ability tests, which tend to favour white males, and could help diversify admissions.

Evidence on inherent SJT cultural biases that may cause these group differentials is mixed. Schaepers et al. (2024) found a German SJT on personal initiative had similar psychometric properties when translated for Cuba, suggesting some SJTs can be adapted cross-culturally. However, Schmitt et al. (2019) analysed cultural content in an SJT for undergraduate admissions, comparing white Americans and Chinese international students. Responses correlated with cultural differences, indicating that cross-cultural – here collectivist and individualist - SJT use requires caution. Additionally, Lievens et al. (2019) found that response formats influence majority-minority results

differentials. Constructed response formats (written or audiovisual) reduce differences compared to multiple-choice formats, likely due to lower cognitive load. The authors note that with technological advances, audiovisual formats may now be an accessible way to minimise cultural biases in SJTs, a need that is increasing with migration flows. **While some SJTs (construct-based, unidimensional) may adapt well across cultures, cultural norms can influence test construction and interpretation**.

A different potential confounder of SJTs is faking. Zhang, Cullen & Sackett (2021) found that **SJTs are susceptible to "faking good",** where individuals inflate their scores by guessing what assessors want. Their study compared SJT performance under low-stakes (946 medical residents) and high-stakes (275 medical residency applicants) conditions, using knowledge- and behaviour-based instructions. Scores were higher in high-stakes scenarios regardless of instruction type, with knowledge-based instructions yielding higher scores overall. However, behaviour-based instructions ("would do" vs "should do") showed greater susceptibility to faking in high-stakes settings. Recent research highlights how **artificial intelligence tools like ChatGPT may amplify this issue**. Harwood, Roulin & Iqbal (2024) observed that SJT scores increased after ChatGPT's release, even in low-stakes mock settings. Detection software struggled to identify AI-generated content, and studies confirmed ChatGPT's proficiency in answering SJTs.

To mitigate faking risks, Xi et al. (2024) explored a "constructed response retest" format, where participants completed SJTs with open-ended responses first, followed by closed responses alongside standard anti-faking warnings. This approach reduced faking further when combined with warnings, likely by increasing the perceived risk of being caught. These findings suggest that adapting assessment methods can help address faking concerns in SJTs.

DISCUSSION

In the findings section, key claims made in the literature were distilled into practice points that indicate the purposes for which each type of psychometric test might be useful. The findings also suggest some broader implications for practitioners in the NSW public sector, five of which are outlined here.

1. **Psychometric tests are most useful as part of a larger, multi-stage selection process.**

Psychometric tests measure various abilities and traits that candidates possess in differing degrees, many of which correlate with work-related outcomes. These abilities and traits are largely distinct but interact, meaning the predictive power of psychometric tests is often additive.

There is limited evidence on the ideal composition or sequence for selection processes. However, given the diversity of constructs measured by psychometric tests, their varying predictive power across roles, and the benefits of contextualising tests within work contexts, their use should be informed by thorough job analysis and capability matching. Tests should align with job requirements, though some have shown moderate predictive power across roles and industries.

This suggests a complex selection process, potentially increasing costs for organisations and candidates. For example, while (structured) interviews and assessment centres are strong predictors of job performance, their costs and scalability challenges make psychometric testing appealing (Griffin, 2018). Conversely, candidates often prefer (unstructured) interviews or open-ended formats, so the use of psychometric tests may attract some candidates but deter others (Schroeder et al, 2025; Van Iddekinge, Lievens & Sackett, 2022).

2. **Psychometric testing shapes the candidate pool and so the potential effects of inherent biases and differential results across groups requires careful monitoring**.

The findings show that overall, the different types of psychometric testing have differential results across societal groups. All else equal, then, increasing the weighting of these tests within a process will influence the composition of the candidate pool (e.g., Woods & Patterson, 2024). In other words, it can amplify actual bias.

Various possibilities for increasing access to candidate pools are available. One approach is to have multiple assessments used in a process (Van Iddekinge, Lievens & Sackett, 2022). Subgroup differences found in cognitive ability testing can be mitigated by adding assessments with smaller subgroup differences. Additionally, pre-test explanations and sample tests can help mitigate racial and socioeconomic disparities in access to coaching.

Opportunity for culturally-aware design, adaptation and use of SJTs for specific roles needs care and deliberate effort, noting the Australian diversity context is unique. The findings from the one study on the UK teaching profession and gender raise a few possible lines of inquiry: teaching is a female-dominated profession in many Anglo-western countries at least. SJTs used to attract women to male dominated areas may not, however, necessarily yield a result if SJTs are shaped by job norms rather than diverse perspectives. Gamification tends to be more attractive to males. Depending on job requirements, considerations like these will be relevant to how psychometric tests are used within larger selection processes.

Some findings indicate that rather than the test, the work or job environment, or other factors, may discriminate based upon race. In other words, a good test result was achievable by minority cohorts but did not match job performance. Strategies to ensure elimination of discrimination in a workplace remain necessary. A challenge with the research is a lack of studies about or on Australia and any unique cultural dynamics. This is relevant especially for First Nations candidates.

The findings from some narrow studies of SJTs also highlight a performance differential depending upon race and culture, favouring white (majority) candidates. These studies, like the tests, are highly occupation specific with two studies on postgraduate student admissions. SJT tests themselves are not uniform.

3. **The use case for non-cognitive psychometric tests is strongest for leadership and managerial roles.**

Non-cognitive psychometric constructs have clear and important implications for predicting work-related outcomes for managerial roles. The Big Five, although generally weak predictors, have also been found to be more predictive for relatively less structured roles, a category that overlaps with management.

Broadly, personality predictive validity is only weak-towards-moderate. The implication is sparing use of personality testing as a main selection tool, and preference among these for conscientiousness testing in general, and integrity tests for specific roles (i.e., customer service and law enforcement).

Overall, predictive validity is generally stronger for EI and SJTs than personality testing, especially traits other than conscientiousness, and these tests should be preferred wherever possible – albeit with caution around the potential biases noted above. Data on key demographics collected at time of testing could be helpful to monitor for gender and culture bias (with usual caveats around a sufficient sample size).

4. The development of artificial intelligence is changing the way that selection operates in practice.

The rapid development of generative artificial intelligence (AI) tools is an emerging issue in selection and recruitment. As noted, AI tools have shown some potential for enabling faking. Additionally, there is early evidence that AI-written answers may improve overall performance in interviews (Canagasuriam & Lukacik, 2025). ChatGPT has shown some ability to perform well on some cognitive ability tests, possibly enabling cheating in unsupervised settings (Hickman, Dunlop & Wolf, 2024). Against this, generative AI might help candidates prepare and, if they are going to be using such tools in their jobs, using them in tests might not reduce validity (Lievens & Dunlop 2024).

For organisations, AI tools are increasingly used to scan resumes and develop position descriptions, among other HR tasks. The broader issues this raises are beyond the scope of this report. However, there are concerns that using AI tools throughout the selection process might reinforce biases. Conversely, AI tools themselves might help address potential biases in selection processes. For example, one recent study found that women candidates' preferred interacting with an AI evaluator than a man, which might be relevant to attracting women to male-dominated fields (Steinberg & Hohenberger, 2023).

Generative AI is a continuation of a trend of technology enabling increased access to and speed in selection processes, traded off with risks around automated reinforcement of bias and destabilisation of performance and behaviour norms by which to assess candidates.

5. Recent literature supports the usefulness of psychometric testing for recruitment and selection – and largely supports current NSW policy.

Current guidance in the NSW public sector supports the use of these 4 types of psychometric testing, which can be chosen to fulfil legal requirements for merit selection processes. This report broadly acknowledges the utility of these tests – especially when used with careful job analysis (provided for in the NSW guide) and in appropriate combinations (NSW Public Service Commission, 2025). The NSW guide also provides advice about the order of the selection process as Table 4 below shows.

Table 5: Order of assessments (from NSW Public Service Commission, 2025)

Assessments	As part of pre-screening	Earlier in process	Later in process
Cognitive ability tests	Yes	Yes	-
Interviews	-	-	Yes
Personality questionnaires	-	-	Yes
Work sample exercises	-	Yes	Yes

This ANZSOG report offers some support for SJTs, which often have a personality loading, to be used in screening candidates. In addition, emotional intelligence testing may be useful to inform selection for management positions.

If a test is administered earlier, steps could be taken to separate or conceal a test ranking where there is a concern about relative test performance influencing selectors in a subsequent selection step.

Finally, Appendix A maps the findings from this ANZSOG report onto the NSW Public Sector Capability Framework (2020), on which job descriptions are based, providing guidance for the composition of selection processes, and reinforcing the utility of these tests.

CONCLUSION AND FUTURE WORK

In conclusion, based on a thematic analysis of empirical evidence to date and noting methodological limitations, the findings are representative of the latest evidence in the field of psychometric testing for selection. On this basis, there are **three headline developments** since NSW OPSC last reviewed the evidence base in 2018:

1. Cognitive ability testing remains the strongest psychometric tool for predicting job performance, but new evidence finds that this relationship is weaker than previously understood and supports the incremental validity of narrow cognitive abilities in some circumstances.

2. Regarding non-cognitive psychometric testing, recent evidence confirms that conscientiousness is the main personality-trait predictor of job performance, while strengthening the importance of this trait and emotional intelligence for management roles and confirming the reliability and moderate predictiveness of SJTs.

3. The academic literature identifies a range of possible biases and confounds affecting psychometric testing, especially how test results are interpreted and incorporated in selection processes, and emergent challenges around faking. Culture is usually demarcated between collectivist and individualist societies, or between black minority and white majority. Australia's context has unique features.

Regarding **future research**, an important gap identified is the relative lack of public sector specific research, especially relevant to Australia. Future work in this area, then, might include governments undertaking studies of their own use of these tests and how they track job performance. Given key findings around the importance of matching selection processes to job descriptions, and unique aspects of public sector employment like integrity requirements, public sector motivation, and hierarchical organisation structures, this research would help confirm fairer and more accurate selection in the public sector, while also contributing to the broader literature.

ΛNZSOG

REFERENCES

Alon-Barkat, S., & Busuioc, M. (2022). Human-AI interactions in public sector decision-making: 'Automation bias' and 'selective adherence' to algorithmic advice. Journal of Public Administration Research &Theory, 33. https://doi.org/10.1093/jopart/muac007

Aşkun, V., Raghavan, M., Ajanovic, E., Çizel, R., & Wiernik, B. M. (2024). Personality and job performance in Türkiye: Psychometric meta-analysis of Turkish studies. Revista de Psicologia del Trabajo y de Las Organizaciones, 40(1), 1-18. https://doi.org/10.5093/jwop2024a1

Bardach, L., Rushby, J. V., & Klassen, R. M. (2021). The selection gap in teacher education: Adverse effects of ethnicity, gender, and socio-economic status on situational judgement test performance. British Journal of Educational Psychology, 91(3), 1015-1034. https://research.ebsco.com/linkprocessor/plink?id=77ae98d8-5927-3568-964e-2b38393cdda5

Beaujean, A. A., & Benson, N. F. (2019). The one and the many: Enduring legacies of Spearman and Thurstone on intelligence test score interpretation. Applied Measurement in Education, 32(3), 198-215. https://research.ebsco.com/linkprocessor/plink?id=4e3175e6-eb37-3ceb-9c7e-a17f36037a72

Bergner, S. (2020). Being smart Is not enough: Personality traits and vocational interests incrementally predict intention, status and success of leaders and entrepreneurs beyond cognitive ability. Frontiers in Psychology, 11.

Berry, C. M., & Zhao, P. (2015). Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance. Journal of Applied Psychology, 100(1), 162-79. https://doi.org/10.1037/a0037615

Bradburn, J., Ryan, A. M., Boyce, A., McKinniss, T., & Way, J. (2020). Is it complicated? Validity of personality interactions for predicting performance. Personnel Assessment & Decisions, 6(3). https://doi.org/10.25035/pad.2020.03.004

Bratton, J., Gold, J., Bratton, A., & Steele, L. (2022). Recruitment and selection. Human Resource Management. Bloomsbury Publishing Plc. https://www.bloomsbury.com/au/human-resource-management-9781352012606/

Canagasuriam, D., & Lukacik, E. R. (2025). ChatGPT, can you take my job interview? Examining artificial intelligence cheating in the asynchronous video interview. International Journal of Selection & Assessment, 33(1). https://doi.org/10.1111/ijsa.12491

Carless, S. A. (2009). Psychological testing for selection purposes: A guide to evidence-based practice for human resource professionals. International Journal of Human Resource Management, 20(12), 2517-2532. https://doi.org/10.1080/09585190903363821

Casillas, A., Robbins, S., McKinniss, T., Postlethwaite, B., & Oh, I. S. (2009). Using narrow facets of an integrity test to predict safety: A test validation study. International Journal of Selection & Assessment, 17(1), 119-125. https://doi.org/10.1111/j.1468-2389.2009.00456.x

Chamorro-Premuzic, T., & Furnham, A. (2010). Situational judgement tests and GPA. In The Psychology of Personnel Selection (pp. 75–92). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819308.005

Christian, M., Edwards, B., & Bradley-Geist, J. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. Personnel Psychology, 63, 83-117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgement tests for selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore & C. Semedo (Eds.), The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention (pp. 75-92). John Wiley & Sons Ltd.

ΛNZSOG

Cunningham, M. L., Douglas, H., & Boag, S. (2018). General mental ability moderates the link between confidence and integrity test scores. Personality & Individual Differences, 123, 94-99. https://doi.org/10.1016/j.paid.2017.11.004

Deary, I. J. (2012). Intelligence. Annual Review of Psychology, 63, 453-482.

Deary, I. J. (2013). Intelligence. Current Biology, 23(16), 673-676.

Doğru, C. (2022). A meta-analysis of the relationships between emotional intelligence and employee outcomes. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.611328

Eshuis, J., de Boer, N., & Klijn, E. H. (2023). Street-level bureaucrats' emotional intelligence and its relation with their performance. Public Administration, 101(3), 804-821. https://doi.org/10.1111/padm.12841

Fisher, D. M., Cunningham, S., Kerr, A. J., & Allscheid, S. P. (2017). Contextualized personality measures in employee selection: Extending frame-of-reference research with job applicant samples. International Journal of Selection & Assessment, 25(1), 18-35. https://doi.org/10.1111/ijsa.12156

Gignac, G. (2010). Seven-factor model of emotional intelligence as measured by Genos EI: A confirmatory factor analytic investigation based on self- and rater-report data. European Journal of Psychological Assessment, 26(4), 309-316. https://doi.org/10.1027/1015-5759/a000041

Gottfredson, L. S. (1994). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Intelligence, 24(1), 13-23.

Government Sector Employment Act 2013 (NSW). https://legislation.nsw.gov.au/view/html/inforce/current/act-2013-040

Government Sector Employment (General) Rules 2014 (NSW). https://legislation.nsw.gov.au/view/html/inforce/current/sl-2014-0564

Griffin, B. (2018). Employee selection: Review of psychometric testing. Macquarie University. Unpublished Report.

Grobelny, J. (2018). The predictive validity toward job performance of general and specific mental abilities: A validity study across different occupational groups. Business and Management Studies, 4(3), 1-12.

Grobelny, J., Radke, P., & Paniotova-Maczka, D. (2021). Emotional intelligence and job performance: A meta-analysis. International Journal of Work Organisation & Emotion, 12, 1-47. https://doi.org/10.1504/IJWOE.2021.10037977

Harenbrock, J., Forthmann, B., & Holling, H. (2023). Retest reliability of situational judgment tests: A meta-analysis. Journal Of Personnel Psychology, 22(4), 169–184. https://doi.org/10.1027/1866-5888/a000323

Harwood, H., Roulin, N., & Iqbal, M. Z. (2024). "Anything you can do, I can do": Examining the use of ChatGPT in situational judgement tests for professional program admission. Journal of Vocational Behaviour, 154, 1-19. https://doi.org/10.1016/j.jvb.2024.104013

He, Y., Donnellan, M. B., & Mendoza, A. M. (2019). Five-factor personality domains and job performance: A second order meta-analysis. Journal of Research in Personality, 82. https://doi.org/10.1016/j.jrp.2019.103848

He, Y., & Gong, R. (2010). A review of integrity tests in personnel selection. In 2010 International Conference on E-Product E-Service and E-Entertainment, E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on (pp. 1-6): IEEE.

Hickman, L., Dunlop, P. D., & Wolf, J. L. (2024). The performance of large language models on quantitative and verbal ability tests: Initial evidence and implications for unproctored high-stakes testing. International Journal of Selection & Assessment, 32(4), 499-511. https://doi.org/10.1111/ijsa.12479

Hughes, D. J., & Batey, M. (2017). Using personality questionnaires for selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention (pp. 151-181). John Wiley & Sons Ltd.

Iliescu, D., Ion, A., Ilie, A., Ispas, D., & Butucescu, A. (2023). The incremental validity of personality over time in predicting job performance, voluntary turnover, and career success in high-stakes contexts: A longitudinal study. Personality and Individual Differences, 213. https://doi.org/10.1016/j.paid.2023.112288

Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the big five personality traits in predicting job performance. The Academy of Management Journal, 58(4), 1149-1179. https://doi.org/10.2307/43589388

Karim, J., & Weisz, R. (2010). Cross-cultural research on the reliability and validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). Cross-Cultural Research: The Journal of Comparative Social Science, 44(4), 374-404. https://doi.org/10.1177/1069397110377603

Kasten, N., Freund, P. A., & Staufenbiel, T. (2020). "Sweet little lies": An in-depth analysis of faking behaviour on situational judgment tests compared to personality questionnaires. European Journal of Psychological Assessment, 36(1), 136-148. https://doi.org/10.1027/1015-5759/a000479

Kepes, S., Keener, S. K., Lievens, F., & McDaniel, M. A. (2024). An integrative, systematic review of the situational judgment test literature. Journal of Management. 1-42. https://doi.org/10.1177/01492063241288545

Lee, H. (2023). Reduction of faking with the use of a forced-choice personality test: Cross-cultural comparisons between South Korea and the United States. International Journal of Selection & Assessment, 31(1), 147-162. https://doi.org/https://doi.org/10.1111/ijsa.12408

Levitats, Z., Vigoda-Gadot, E., & Vashdi, D. R. (2019). Engage them through emotions: Exploring the role of emotional intelligence in public engagement. Public Administration Review, 79(6), 841-852. https://doi.org/10.1111/puar.13113

Leyens, T., Ritz, J., & Hofmans, J. (2024). Personality. In M. Bal (Ed.), Encyclopedia of organizational psychology (pp. 512–517). Edward Elgar. https://doi.org/10.4337/9781803921761.00101

Lievens, F. (2005). International situational judgement tests. In Weekley, J. A., & Ployhart, R. E. (Eds.). (2006). Situational Judgment Tests: Theory, Measurement, and Application (pp. 279-300). Taylor & Francis Group. https://research.ebsco.com/linkprocessor/plink?id=b5251939-9225-3d7d-b0ec-d4b475dd9b16

Lievens, F., & Chan, D. (2017). Practical intelligence, emotional intelligence, and social intelligence. In J. L. Farr, & N. T. Tippins (Eds.), Handbook of Employee Selection (pp. 342-364). Routledge. https://doi.org/10.4324/9781315690193-15

Lievens, F., & Dunlop, P. D. (2025). Effects of applicants' use of generative AI in personnel selection: Towards a more nuanced view? International Journal of Selection & Assessment, 33(1). https://doi.org/10.1111/ijsa.12516

Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: Evidence from the UKCAT. Medical Education, 50(6), 624-636. https://doi.org/10.1111/medu.13060

Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & Soete, B. D. (2019). Constructed response formats and their effects on minority-majority differences and validity. Journal of Applied Psychology, 104(5), 715-726. https://doi.org/10.1037/apl0000367

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behaviour: Big five, or one in addition? Personnel Psychology, 60(1), 1-34. https://doi.org/10.1111/j.1744-6570.2007.00063.x

McCrae, R. R., Costa Jr, P. T., McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. Journal of Personality, 57(1), 17-40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. Personnel Psychology, 60(1), 63–91. https://doi.org/10.1111/j.1744-6570.2007.00065.x

Miao, C., Humphrey, R. H., & Qian, S. (2016). A meta-analysis of emotional intelligence and work attitudes. Journal of Occupational & Organizational Psychology, 90(2), 177-202. https://doi.org/10.1111/joop.12167

Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive ability and job performance: Meta-analytic evidence for the validity of narrow cognitive abilities. Journal of Business & Psychology, 37(6), 1119-1139. https://doi.org/10.1007/s10869-022-09796-1

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology, 78(4), 679-703. https://doi.org/10.1037/0021-9010.78.4.679

Palmer, B. R., Stough, C., Harmer, R., & Gignac, G. (2009). The Genos Emotional Intelligence Inventory: A measure designed specifically for workplace applications. In J. D. A. Parker, D. H. Saklofske, & C. Stough (Eds.), Assessing Emotional Intelligence: Theory, Research, and Applications (pp. 103-117). Springer US. https://doi.org/10.1007/978-0-387-88370-0_6

Patterson, F., Rowett, E., Hale, R., Grant, M., Roberts, C., Cousans, F., & Martin, S. (2016). The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia. BMC medical education, 16, 87. https://doi.org/10.1186/s12909-016-0606-4

Pereira Teques, A. P. de S, Cavalcanti, H. A. F., Barbosa, L. N. F., & Teques, P. H. A. (2023). Cross-cultural adaptation and validation of the emotional intelligence self-perception questionnaire. Paideia (0103863X), 33, 1–9. https://doi.org/10.1590/1982-4327e3316

Pittenger, D. (1993). Measuring the MBTI … and coming up short. Journal of Career Planning and Employment, 54.

Potočnik, K., Anderson, N. R., Born, M., Kleinmann, M., & Nikolaou, I. (2021). Paving the way for research in recruitment and selection: recent developments, challenges and future opportunities. European Journal of Work & Organizational Psychology, 30(2), 159-174. https://doi.org/10.1080/1359432X.2021.1904898

Public Service Commission (2025). Recruitment and Selection Guide. https://www.psc.nsw.gov.au/workforce-management/recruitment/recruitment-and-selection-guide (accessed 19 March 2025)

Public Service Commission (2020). The NSW Public Sector Capability Framework Version 2. https://www.psc.nsw.gov.au/assets/psc/documents/capability_framework_v2_2020.pdf (accessed 20 March 2025)

Randall, K., Isaacson, M., & Ciro, C. (2017). Validity and reliability of the Myers-Briggs Personality Type Indicator: A systematic review and meta-analysis. Journal of Best Practices in Health Professions Diversity, 10(1), 1-27. https://research.ebsco.com/linkprocessor/plink?id=368850fa-551b-39de-a35c-89dc00286516

Rivers, S. E., Handley-Miner, I. J., Mayer, J. D., & Caruso, D. R. (2020). Emotional intelligence. In R. J. Sternberg (Ed.), The Cambridge Handbook of Intelligence (2 ed., pp. 709-735). Cambridge University Press. https://doi.org/10.1017/9781108770422.030

Roberts, B., Walton, K., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin, 132, 1-25. https://doi.org/10.1037/0033-2909.132.1.1

Robinson, M. D. (2024). Ability-related emotional intelligence: An introduction. Journal of Intelligence, 12(5). https://doi.org/10.3390/jintelligence12050051

Rosales, V., Conley, C., & Norris, M. C. (2024). Racial and ethnic disparities in situational judgment testing among applicants to an anesthesiology residency program. Journal of Graduate Medical Education, 16(2), 140-145.

Rust, J., Golombok, S., Kosinski, M., & Golombok, S. (2009). Modern Psychometrics: The Science of Psychological Assessment. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=1734156

Sackett, P. R., Demeke, S., Bazian, I. M., Griebie, A. M., Priest, R., & Kuncel, N. R. (2024). A contemporary look at the relationship between general cognitive ability and job performance. Journal of Applied Psychology, 109(5), 687-713. https://doi.org/10.1037/apl0001159

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. Journal of Applied Psychology, 102(3), 254-273. https://doi.org/10.1037/apl0000151

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. Journal of Applied Psychology, 107(11), 2040-2068. https://doi.org/10.1037/apl0000994

Salgado, J. F. (2017). Moderator effects of job complexity on the validity of forced-choice personality inventories for predicting job performance. Revista de Psicologia del Trabajo y de Las Organizaciones, 33(3), 229-238. https://doi.org/10.1016/j.rpto.2017.07.001

Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. Journal of Occupational Organizational Psychology, 88(4), 797–834. https://doi.org/10.1111/joop.12098

Salgado, J. F., & Moscoso, S. (2019). Meta-analysis of the validity of general mental ability for five performance criteria: Hunter and Hunter (1984) revisited. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.0222

Schaepers, P., Heinemann, H., Alvarez, D. P., Nohr, L., & Monke, F. W. (2024). Different culture, same situation? Translating and applying a situational judgment test from Germany in Cuba. Psychological Test Adaptation & Development, 5, 1-11.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124(2), 262. https://research.ebsco.com/linkprocessor/plink?id=b67b980b-62cd-3d69-b9f1-ceb24539bc78

Schmitt, N., Prasad, J. J., Ryan, A. M., Bradburn, J. C., & Nye, C. D. (2019). Culture as a determinant of option choice in a situational judgement test: A new look. Journal of Occupational & Organizational Psychology, 92(2), 330-351. https://doi.org/10.1111/joop.12251

Schroeder, V. S., Kleinmann, M., Heimann, A. L., & Ingold, P. V. (2025). Why do applicants (dis)Like selection methods? The role of stimulus and response format for need satisfaction. Journal of Business & Psychology, 1-22. https://doi.org/10.1007/s10869-024-10002-7

Sosnowska, J., Hofmans, J., & Lievens, F. (2018). Assessing personality dynamics in personnel selection. In J. F. Rauthmann (Ed.), The Handbook of Personality Dynamics and Processes (pp. 1139-1157). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-813995-0.00044-3

Steinberg, S. A. L., & Hohenberger, C. (2023). Can AI close the gender gap in the job market? Individuals' preferences for AI evaluations. Computers in Human Behaviour Reports, 10. https://doi.org/10.1016/j.chbr.2023.100287

Te Nijenhuis, J., Pesta, B. J., & Fuerst, J. G. R. (2024). General mental ability testing and adverse impact in the United Kingdom: A meta-analysis with more than two million observations. European Journal of Work & Organizational Psychology, 33(5), 712-725. https://doi.org/10.1080/1359432X.2024.2377780

Tiffin, P. A., Paton, L. W., O'Mara, D., MacCann, C., Lang, J. W. B., & Lievens, F. (2019). Situational judgement tests for selection: Traditional vs construct-driven approaches. Medical Education, 54(2), 105-115. https://doi.org/10.1111/medu.14011

Triyono, T., & Tjahjono, H. K. (2023). The effect of emotional intelligence on public sector employee's affective commitment. Journal of Governance & Public Policy, 10(2), 118–128. https://doi.org/10.18196/jgpp.v10i1.14007

Van Iddekinge, C. H., Lievens, F., & Sackett, P. R. (2023). Personnel selection: A review of ways to maximize validity, diversity, and the applicant experience. Personnel Psychology, 76(2), 651-686. https://doi.org/10.1111/peps.12578

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. Journal of Applied Psychology, 97(3), 499-530. https://doi.org/10.1037/a0021196

van Lill, X., Stols, A., Rajab, P., & Wiggett, J. (2023). The validity of a general factor of emotional intelligence in the South African context. In African Journal of Psychological Assessment (Vol. 5, pp. e1-e10): AOSIS.

van Zyl, C. J. J. (2014). The psychometric properties of the Emotional Quotient Inventory 2.0 in South Africa. SAJIP: South African Journal of Industrial Psychology, 40(1), 1-8. https://doi.org/10.4102/sajip.v40i1.1192

Webster, E. S., Paton, L. W., Crampton, P. E. S., & Tiffin, P. A. (2020). Situational judgement test validity for selection: A systematic review and meta-analysis. Medical Education, 54(10), 888-902. https://doi.org/10.1111/medu.14201

Whetzel, D. L., Sullivan, T. S., & McCloy, R. A. (2020). Situational judgment tests: An overview of development practices and psychometric characteristics. Personnel Assessments & Decisions, 6(1).

Wing-Man Lau, R., Chan, D. K. S., Sun, F., & Cheng, G. H. L. (2023). Predictive validity of integrity tests for workplace deviance across industries and countries in the past 50 years: A meta-analytic review. Journal of Pacific Rim Psychology, 17. https://doi.org/10.1177/18344909231171729

Wonda, T. A. (2024). Does emotional intelligence have a contribution to employee performance enhancement? Empirical evidence from the public sectors. Cogent Business & Management, 11(1), 2415525. https://doi.org/10.1080/23311975.2024.2415525

Woods, S. A., & Patterson, F. (2024). A critical review of the use of cognitive ability testing for selection into graduate and higher professional occupations. Journal of Occupational & Organizational Psychology, 97(1), 253-272. https://doi.org/10.1111/joop.12470

Xi, L., Weng, Q., Wang, X., Chen, L., & Corstjens, J. (2024). Effects of a constructed response retest strategy on faking, test perceptions, and criterion-related validity of situational judgment tests. International Journal of Selection & Assessment, 32(4), 561-578. https://doi.org/10.1111/ijsa.12482

Zell, E., & Lesick, T. L. (2022). Big five personality traits and performance: A quantitative synthesis of 50+ meta-analyses. Journal of Personality, 90(4), 559-573. https://doi.org/10.1111/jopy.12683

Zhang, C., Cullen, M. J., & Sackett, P. R. (2021). Effects of response instructions on situational judgment test performance in operational selection and developmental contexts. International Journal of Selection & Assessment, 29(2), 170-184. https://doi.org/10.1111/ijsa.12324

APPENDIX A

Table 6: Mapping key findings to the NSW Public Sector Capability Framework

Capability	Relevant findings	Reference
Display resilience and courage	Openness and extraversion as measured by Big Five personality tests are somewhat predictive of job performance. The effect is stronger for customer service jobs. Emotional intelligence is correlated with lower likelihood of experiencing job stress.	Zell & Lesick, 2022; He, Donnellan & Mendoza, 2019; Doğru, 2022.
Act with integrity	General cognitive ability is correlated with organisational citizenship behaviours, as is emotional intelligence. This latter effect is stronger for non-managerial employees. Integrity tests predict counterproductive behaviours and are moderate to strong predictors of job performance.	Nye et al, 2022; Doğru, 2022; Wing-Man Lau et al, 2023; Ones et al, 1993.
Manage self	Conscientiousness is the personality trait most strongly predictive of job performance – and adds validity over and above cognitive ability.	Zell & Lesick, 2022; Iliescu et al, 2023.
Value diversity and inclusion	Agreeableness is somewhat predictive of job performance and adds incremental validity over and above cognitive ability.	Zell & Lesick, 2022; Iliescu et al, 2023.
Communicate effectively	General intelligence includes verbal reasoning and is a predictor of job performance. Extraversion and openness are also both correlated with job performance.	Sackett et al, 2024; Zell & Lesick, 2022.
Commit to customer service	Emotional intelligence is most strongly correlated with job performance in call centres, and it is correlated with job performance for street-level bureaucrats.	Grobelny, Radke & Paniotova-Macza, 2021; Eshuis, de Boer & Klijn, 2021.
Work collaboratively	Situational judgement tests have been found to be moderately predictive of teamwork.	Christian et al, 2010.
Influence and negotiate	Emotional intelligence includes recognising emotions in others and is predictive of job performance.	Bergner, 2020
Deliver results	General cognitive ability is a moderate-to-strong predictor of job performance across occupations and outcomes.	Sackett et al, 2024; Salgado & Moscoso, 2019; Nye et al, 2022.
Plan and prioritise	Situational judgement tests are moderate predictors of job performance and add validity over and above cognitive ability and personality.	Webster et al, 2020; McDaniel et al, 2007.
Think and solve problems	General cognitive ability includes the ability to apply learning. Behaviour-based situational judgement tests measure how people respond to scenarios and have incremental validity over general cognitive ability.	Sackett et al, 2024; McDaniel et al, 2007.
Demonstrate accountability	Emotional intelligence is correlated with organisational commitment and organisational citizenship behaviours.	Doğru, 2022
Finance	Cognitive ability, including verbal and numerical reasoning, may be relevant here as a moderate predictor of job performance.	Sackett et al, 2024.

ΛΝZSOG

Capability	Relevant findings	Reference
Technology	Narrow cognitive abilities, including visual-spatial reasoning, can add predictive validity over and above general intelligence.	Nye et al, 2022.
Procurement and contract management	Knowledge-based situational judgement tests measure how people apply rules in different scenarios and are moderate predictors of job performance.	McDaniel et al, 2007.
Project management	Conscientiousness is correlated with task performance.	Zell & Lesick, 2022.
Manage and develop people	Cognitive ability is correlated with leadership success, moderated by openness and emotional stability. Cognitive ability is slightly *less* predictive of managers' job performance than non-managers; whereas the relationship is possibly reversed for emotional intelligence.	Bergner, 2020; Sackett et al, 2024; Doğru, 2022 (but cf Grobelny, Radke & Paniotova-Macza, 2021).
Inspire direction and purpose	Extraversion is correlated with leadership success, as are, to a lesser degree, openness and emotional stability.	Bergner, 2022.
Optimise business outcomes	Managers' emotional intelligence influences the connection between staff emotional intelligence and job performance.	Levitats, 2019.
Manage reform and change	Emotional intelligence is correlated with managers' organisational commitment.	Dogru, 2022; Grobelny, Radke & Paniotova-Macza, 2021.
Specialised capabilities	Narrow cognitive abilities can add predictive validity over and above general intelligence. Contextualising personality tests in the work context can increase their predictiveness.	Nye et al, 2022; Fisher et al, 2017.

Based on *NSW Public Sector Capability Framework 2020*, available via:

https://www.psc.nsw.gov.au/assets/psc/documents/capability_framework_v2_2020.pdf