Work-Contextualized Situational Judgment Tests: Evaluating Reliability and Validity Evidence for Use with High School Students

Fernando Mena Serrano, Kate E. Walton, Dana Murano, and Jeremy Burrus

Abstract

We evaluated the reliability and validity evidence of work-based (i.e., scenarios that would be expected in a typical workplace) situational judgment tests (SJTs) relative to that of school-based SJTs in two samples of high school students. In Study 1, participants (*N* = 188) responded to SJTs assessing Conscientiousness, Agreeableness, and Emotional Stability. In Study 2, participants (*N* = 233) responded to SJTs assessing Openness, Extraversion, and Integrity. We also administered the Big Five Inventory to test for convergent/discriminant validity and the Job Performance Scale to test for criterion validity. Work-based SJTs generally demonstrated acceptable psychometric properties, though the work-based Conscientiousness and Integrity scales had significantly lower reliability than school-based versions, and Extraversion showed poor internal consistency in both contexts. Conscientiousness consistently predicted in-role behavior and correlated with organizational citizenship behavior, while Agreeableness most strongly predicted organizational citizenship in the work-based format. Overall, findings support the feasibility of work-based SJTs in high school settings.

Introduction

The ACT® WorkKeys® Essential Skills (WKES) assessment is a workplace-focused measure designed to evaluate six essential skills: Work Ethic, Collaboration, Resilience, Creativity, Leadership, and Integrity. The first five domains align with the factors from the Big Five personality model (Digman, 1990), while Integrity aligns with the Honesty-Humility factor of the HEXACO model (Ashton & Lee, 2007). Table 1 shows the crosswalk between the WKES essential skills and Big Five/HEXACO factor. Each skill is assessed through both self-report Likert-type items and SJTs (described in more detail immediately below), which are combined into an aggregate domain score. Intended for use with current and prospective employees, the assessment supports applications in employee development, coaching, and selection[1]. Initial validation studies with employed adults provide strong reliability evidence (Mosier reliability estimates between .83 and .91) and support for construct validity, showing expected correlations with corresponding Big Five and HEXACO traits, as well as criterion validity, with significant associations between certain skills (e.g., Work Ethic, Collaboration, Creativity) and supervisor-rated job performance (ACT, 2024).

[1] If used for selection, we strongly recommend it be used in conjunction with other indicators of work readiness and job fit, such as interviews and other assessments.

Table 1. Crosswalk Between WKES Essential Skills and Big Five/HEXACO Factors

WKES essential skill	Definition	Big Five/HEXACO factor (and its definition)
Work Ethic	The extent to which a person's actions demonstrate persistence, goal striving, reliability, dependability, and attention to detail at work	Conscientiousness (The tendency to be organized, dependable, and hardworking while controlling impulses to achieve goals)
Resilience	The extent to which a person's actions demonstrate stress management, emotional regulation, a positive response to setbacks, and poise	Emotional Stability (The ability to stay composed and regulate negative emotions under stress or uncertainty)
Collaboration	The extent to which a person's actions demonstrate the ability to work on teams, empathy, helpfulness, trust, and trustworthiness	Agreeableness (A prosocial orientation marked by being cooperative, tolerant, friendly, and trustworthy)
Leadership	The extent to which a person's actions demonstrate assertiveness, influence, optimism, and enthusiasm	Extraversion (A desire for active social interaction, involving sociability, assertiveness, enthusiasm, and influence)
Creativity	The extent to which a person's actions demonstrate ingenuity, creative thinking, inquisitiveness, flexibility, open-mindedness, and embracing diversity	Openness (A curiosity and acceptance of the unfamiliar, characterized by creativity, thoughtfulness, and open-mindedness)
Integrity	The extent to which a person's actions demonstrate honesty, sincerity, fairness toward others, and modesty at work	Honesty-Humility (The tendency to be sincere, fair, and modest, avoiding manipulation, greed, or entitlement in dealings with others)

SJTs, one of the two item types used in WKES, present examinees with brief, context-rich scenarios and several plausible courses of action. Examinees are asked to judge which responses are more or less effective (Weekley & Ployhart, 2006) or are instructed to indicate how likely they would be to engage in each course of action. WKES employs the latter format. SJTs are useful for assessing behavioral and social and emotional competencies because they sample judgment in realistic situations, minimize construct-irrelevant demands (when appropriately framed), and can target the same latent traits across multiple contexts (Christian et al., 2010).

With the current studies, our objective was to evaluate whether work-based SJTs remain reliable and valid when completed by high school students who may not have the work experience of a typical adult. Specifically, we compare reliability and validity estimates of work-based SJTs with estimates of school-based SJTs, which we use as a benchmark, assuming the school-based scenarios are familiar to high school students and therefore should be reliable and valid. The scenarios provided are not specific to the WKES product and therefore should

generalize to other school- and work-contextualized SJT assessments. For this reason, we use the Big Five/HEXACO terminology rather than the WKES skill labels for the remainder of the report.

The research questions guiding this study are:

1. Is there a difference in reliability between school- and work-based SJTs? Does it vary by skill?
2. Is there a difference in convergent, discriminant, or test-criterion validity estimates between school- and work-based SJTs? Does it vary by skill?
3. Do people report the same level of confidence when responding to school- and work-based SJTs? Does it vary by skill?
4. How do students with no work experience respond to work-based SJTs?

Method

To decrease the length of the study and the burden on participants, the six skills were divided into two studies: Study 1 included SJTs related to Conscientiousness, Agreeableness, and Emotional Stability, and Study 2 focused on Openness, Extraversion, and Integrity. Each study included work-based and school-based SJTs.

Participants

In Study 1, we surveyed 188 11th-12th graders in two waves (101 participated between June 14 and June 23, 2025, and 87 participated between July 12 and July 22, 2025). Participants were students who took the ACT® National test on the June or July 2025 national test dates. After the June test date, an invitation went out to a random sample of 4,225 students inviting them to participate in research. After the July test date, an invitation went out to a random sample of 5,000 students. As noted, we have complete data for 188 students, which is 2.0% of the invited students.[2]

In Study 2, we surveyed 233 11th-12th graders between July 13 and July 15, 2025. These were students who took the ACT® National test on the July 2025 national test date. The invitation to participate went out to 10,000 students. Complete data are available for 2.3% of the invited students.

Students were not incentivized to participate, and they were assured that their involvement and responses would not impact their ACT scores. Demographic data (specifically, gender and race/ethnicity) were available through ACT registration files, and a summary of these can be found in Table 2. This table also shows responses to a question about the maximum hours students have worked in a typical week.

[2] This is the typical response rate to ACT's post-ACT survey invitations.

Table 2. Demographic Characteristics by Study

Characteristic	Level	Study 1 (*N* = 188)	Study 2 (*N* = 233)
Gender	Female	73.9%	72.5%
	Male	26.1%	27.5%
Race/ethnicity	White	58.5%	58.4%
	Hispanic/Latino	11.7%	11.2%
	Asian	10.6%	12.4%
	Black/African American	6.9%	8.6%
	Multiracial	6.9%	3.0%
	Undisclosed	4.8%	6.0%
	American Indian/ Alaska Native	0.5%	0.4%
Work exposure (hours/week)	>20	29.8%	32.2%
	10–20	30.9%	19.7%
	1–10	13.8%	16.3%
	0	25.5%	31.8%

Measures

The SJTs used in this study were not part of WKES, although they followed a similar structure and response format. They were designed so that the school- and work-based versions were nearly identical, differing only in contextual details. Table 3 presents an example of a school-based and a work-based SJT and illustrates how they differed from one another, with differences marked by brackets [ ]. Respondents indicate how likely they are to engage in each of the actions on a five-point scale including the following response options: very unlikely, unlikely, may or may not, likely, and very likely.

Table 3. Example of School- and Work-Based SJTs

School-based SJT	Work-based SJT
You need to revise [your English essay] before handing it in, using a checklist that has been provided. The checklist identifies potential errors and provides space for you to check off each item after review. The checklist and revised [essay] are due tomorrow. How likely are you to do each of the following? • Read through the [essay] and the checklist, then turn them in. • Proofread the [essay] in your typical manner then go through it again with the checklist. • Turn in the [essay] without proofreading it.	You need to revise [the report your boss is expecting] before handing it in, using a checklist that has been provided. The checklist identifies potential errors and provides space for you to check off each item after review. The checklist and revised [report] are due tomorrow. How likely are you to do each of the following? • Read through the [report] and the checklist, then turn them in. • Proofread the [report] in your typical manner then go through it again with the checklist. • Turn in the [report] without proofreading it.

Note. Text underlined and in brackets represents the elements that differ between versions.

We examined the convergent and discriminant validity of the school- and work-based SJTs using the 15-item Big Five Inventory-2-Extra Short Form (BFI-2-XS; Soto & John, 2017), along with three additional items from the Honesty-Humility scale of Ashton and Lee's (2009) HEXACO Personality Inventory-Revised. We examined test-criterion validity using a revised version of the Job Performance Scale (JPS), specifically the In-Role Behavior (IRB) and Organizational Citizenship Behavior (OCB) subscales (Williams & Anderson, 1991). The JPS was revised in three ways. First, only the five items from the IRB and OCB subscales with the highest factor loadings were used. Second, the items were revised to be written in the first person. Third, one item about phone usage was altered to reflect modern-day behaviors. The BFI-2-XS and JPS are Likert-type measures with response options ranging from 1 (strongly disagree) to 6 (strongly agree). The students were asked to rate their confidence in their responses to each SJT on a five-point Likert-type scale (not at all, slightly, somewhat, quite, or extremely confident).

Students who indicated they had no prior work experience were asked an open-ended question that read, "How were you able to respond to the above work-based situations?" In addition, they were asked the question, "Given that you have never been employed, how confident do you feel answering questions about how you would behave in the workforce?" They responded on the same five-point confidence scale described above.

Results

Study 1 (Conscientiousness, Agreeableness, and Emotional Stability)

Reliability of the School- and Work-Based SJTs

Internal consistency reliability (α) was estimated for the school- and work-based SJTs for each trait, and within-sample comparisons of α employed Feldt's t-tests (Feldt, 1980) for two dependent coefficients. The results in Table 4 show that no significant differences were found between the reliability coefficients of the Agreeableness scales or the Emotional Stability scales on the sample of 188 11th-12th graders. By contrast, the Conscientiousness scales differed significantly, with the school-based SJT displaying the higher reliability coefficient (α = .68 vs. .57).

Table 4. Feldt's *t*-test Comparing the Reliability Alpha Coefficients of the School- and Work-Based SJTs (Study 1)

Factor	School-based SJT	Work-based SJT	*p* value
Conscientiousness	.68	.57	.01*
Agreeableness	.52	.58	.19
Emotional Stability	.62	.68	.07

Note. * indicates statistically significant difference ($p < .05$).

Convergent and Discriminant Validity of the School- and Work-Based SJTs

The correlations between the SJTs and the Big Five/HEXACO factors support the school-based SJTs' convergent and discriminant validity. Convergent evidence is clear for Agreeableness (r = .60), Emotional Stability (r = .45), and Conscientiousness (r =.43), with much higher correlations than the off-trait relationships. The mean off-trait correlations, an index of discriminant validity, ranged from .19 for Emotional Stability to .31 for Agreeableness.

The work-based SJTs also demonstrated convergent validity. Conscientiousness, Agreeableness, and Emotional Stability showed their highest correlations with their parallel BFI dimensions (r = .35, .46, and .49, respectively), as shown in Table 5. Also, Conscientiousness correlated almost as strongly with SJT Openness (r = .32) as with its target trait. The mean off-trait correlations ranged from .15 for Emotional Stability to .23 for Agreeableness.

Table 5. Pearson Correlation Coefficients Between the School- and Work-Based SJT Scales and the Big Five/HEXACO Subscales (Study 1)

Big Five/HEXACO	School-based SJT			Work-based SJT		
	C	A	ES	C	A	ES
Conscientiousness	.43*	.34*	.35*	.35*	.34*	.32*
Agreeableness	.32*	.60*	.17*	.22*	.46*	.15
Emotional Stability	.05	.07	.45*	–.04	.16	.49*
Openness	.31*	.40*	.17*	.32*	.36*	.05
Extraversion	.26*	.26*	.23*	.11	.18*	.20*
Integrity	.29*	.48*	.05	.21*	.35*	.04
Mean off-trait correlation	.25	.31	.19	.17	.23	.15

Note. * = p < .05, C = Conscientiousness, A = Agreeableness, ES = Emotional Stability. In-trait correlations (indices of convergent validity) are underlined.

We carried out statistical comparisons of the magnitudes of the correlations presented in Table 5 to determine whether the school- or work-based SJTs had stronger convergent estimates (Williams, 1959; Table 6). For example, we compared the school SJT-BFI Conscientiousness correlation (r = .43) with the work SJT-BFI Conscientiousness correlation (r = .35). With respect to convergent validity, the Agreeableness correlations differed significantly, [t(141) = 2.92, p < .01], indicating a stronger association in the school setting (r = .32) than in the work setting (r = .22).

Table 6. Williams's *t*-test Comparing the Correlation Coefficients Between the School- and Work-Based SJTs Scales and the Big Five/HEXACO Subscales (Study 1)

School- and work-based comparison	Williams's *t*-test (df)	*p* value
Conscientiousness → Big Five/HEXACO Conscientiousness	1.28 (154)	.20
Agreeableness → Big Five/HEXACO Agreeableness	2.92 (141)	<.01*
Emotional Stability → Big Five/HEXACO Emotional Stability	.76 (133)	.45

Note. * indicates statistically significant differences ($p < .05$).

Test-Criterion Validity of the School- and Work-Based SJTs

Across both SJT versions (see Table 7), Conscientiousness and Agreeableness displayed small-to-moderate, significant associations with IRB ($r = .34–.52$) and OCB ($r = .28–.48$). By contrast, Emotional Stability showed non-significant correlations with both subscales. To get a sense of what is typically found in the literature, we also included in Table 7 findings from a meta-analysis (Pletzer et al., 2021), which examines the relationship between HEXACO personality factors and OCB. As Pletzer et al. found, Conscientiousness was the factor most highly related to overall job performances, and both the school- and work-based SJTs outperformed the average effect reported in the meta-analysis, as did the Agreeableness SJTs. The small effects found for Emotional Stability are largely on par with what Pletzer et al. reported.

Table 7. Pearson Correlation Coefficients Between the School- and Work-based SJT Scales and the JPS Subscales (Study 1)

SJT scale	In-role behavior		Organizational citizenship behavior		Pletzer et al. (2021) meta-analysis
	School-based	Work-based	School-based	Work-based	
Conscientiousness	.52*	.47*	.48*	.41*	.32
Agreeableness	.39*	.34*	.45*	.28*	.22
Emotional Stability	.04	.07	.13	.15	<.01†

Note. * = $p < .05$. † = absolute value reported, as the study conceptualized it as Neuroticism, the opposite of Emotional Stability.

We carried out statistical comparisons of the magnitudes of the correlations presented in Table 7 to determine whether the school- or work-based SJTs had stronger test-criterion estimates (see Table 8). The school-based Agreeableness SJT scale had a stronger association with OCB ($r = .45$) than the work-based Agreeableness SJT scale ($r = .28$) did [$t(141) = 3.10$, $p < .01$].

Table 8. Williams' *t*-Test Comparing the Correlation Coefficients Between the School- and Work-Based SJTs and JPS Subscales (Study 1)

School- and work-based comparison	Williams's *t*-test (df)	*p* value
Conscientiousness → IRB	.78 (154)	.44
Agreeableness → IRB	1.01 (141)	.31
Emotional Stability → IRB	.42 (133)	.67
Conscientiousness → OCB	.44 (154)	.66
Agreeableness → OCB	3.10 (141)	<.01
Emotional Stability → OCB	.18 (133)	.85

Regression analyses were conducted to examine the predictive validity of the school-based and work-based SJT scales with IRB and OCB subscales as the outcome variables. When taken as a whole, the three school-based SJT scales accounted for 28% of the variance of the IRB scale (R^2 = .28), whereas the three work-based SJT scales accounted for 25% of its variance (R^2 = .25). When considered jointly, the three school-based SJTs explained more than 27% of the variance of the OCB scale (R^2 = .27), while the three work-based SJTs explained just 19% of the variance of the OCB scale (R^2 = .19). Complete model statistics can be found in the Appendix.

Level of Confidence Across School- and Work-Based SJTs

Recall that we asked students how confident they felt when responding to the SJTs. Across both SJT contexts, students' mean confidence rating were high, as at least 73% of respondents described themselves as "quite" or "extremely" confident in their ratings (Table 9). Note that more students said they were quite or extremely confident in their ratings when responding to work-contextualized scenarios vs. school-contextualized scenarios.

Table 9. Mean Confidence Levels of Students in Their Ratings of School and Work-Based SJTs Subscales (Study 1)

Context	SJT scale	Mean confidence in ratings	
		Not at all, slightly, or somewhat confident	Quite or extremely confident
School-based	Conscientiousness	20.1%	79.9%
	Agreeableness	22.0%	78.0%
	Emotional Stability	26.1%	73.9%
Work-based	Conscientiousness	16.4%	83.6%
	Agreeableness	15.9%	84.1%
	Emotional Stability	24.5%	75.5%

Responses of Students with no Work Experience

Also recall we asked the students with no work experience how confident they feel answering questions about how they would behave in the workforce on a five-point scale. More than one-third (39.0%) reported feeling only somewhat confident when judging how they would behave on the job, whereas nearly half (46.3%) reported feeling quite confident, and the rest (14.6%) reported feeling extremely confident.

We analyzed 33 open-ended responses to the question about how the students with no work experience responded to the work-based SJTs. Most of these students said they answered the work-based SJT by mapping the scenarios to familiar settings (e.g., school projects, group assignments, theatre productions, or sports teams) where they already practice deadlines, cooperation, and leadership. Several mentioned that the work items closely mirrored the school-based ones, so they reused the same reasoning. When direct parallels were lacking, students imagined how their personality, morals, and general work ethics would translate to a job.

Study 2 (Openness, Extraversion, and Integrity)

Study 2 was designed to evaluate the research questions as they pertain to the three final skills aligning to the Big Five and HEXACO frameworks—Openness, Extraversion, and Integrity. Recall, the sample consisted of 223 students in grades 11–12.

Reliability of the School- and Work-Based SJTs

The results in Table 10 show that no significant differences were found in the reliability coefficients between the Openness scales or the Extraversion scales. By contrast, the Integrity scales differed significantly, with the school-based SJTs displaying a higher reliability coefficient (α = .75 vs. .58).

Table 10. Feldt's *t*-test Comparing the Reliability Alpha Coefficients of the School- and Work-Based SJTs (Study 2)

Factor	School-based SJT	Work-based SJT	*p* value
Openness	.58	.57	.87
Extraversion	.34	.38	.46
Integrity	.75	.58	<.01

Extraversion's low reliability indices warranted a more in-depth analysis. In both assessments, a specific item showed poor internal consistency, with low or even negative item-total correlations. Given these findings, we decided not to proceed with validity analyses for the Extraversion factor. The subsequent analyses focus only on the Openness and Integrity scales.

Convergent and Discriminant Validity of the School- and Work-Based SJTs

The correlations between the SJTs and HEXACO in Table 11 support the school-based SJT's convergent and discriminant validity. Convergent evidence is clear for Openness (r = .44) and Integrity (r = .60), with higher correlations than most of the off-trait relationships, although the Openness SJT and Extraversion correlation (r = .48) was higher than the in-trait relation.

The work-based SJT also demonstrated convergent validity in a similar pattern. Openness and Integrity showed their highest correlations with the parallel SJT dimensions (r = .37 and .61, respectively). The off-trait dimensions also showed low correlations, adding evidence of discriminant validity.

Table 11. Pearson Correlation Coefficients Between the School- and Work-Based SJT Scales and the Big Five/HEXACO Subscales (Study 2)

Big Five/HEXACO	School-based SJT		Work-based SJT	
	Openness	Integrity	Openness	Integrity
Conscientiousness	.32*	.31*	.21*	.18*
Agreeableness	.28*	.34*	.25*	.25*
Emotional Stability	.12	.14	.08	.21*
Openness	.44*	.26*	.37*	.18*
Extraversion	.48*	.08	.28*	.13
Integrity	.22*	.60*	.14	.61*
Mean off-trait correlation	.28	.23	.21	.19

Note. * = p < .05. In-trait correlations (indices of convergent validity) are underlined.

A comparison of the magnitudes of the school- and work-based SJT correlations with the Big Five/HEXACO subscales in Table 12 revealed that, with respect to convergent validity, neither Openness nor Integrity differed significantly.

Table 12. Williams's *t*-test Comparing the Correlation Coefficients Between the School- and Work-Based SJTs and the Big Five/HEXACO Subscales (Study 2)

School- and work-based comparison	Williams's *t*-test (df)	*p* value
Openness → Big Five/HEXACO Openness	1.54 (117)	.12
Integrity → Big Five/HEXACO Integrity	.044 (129)	.97

Also, when it comes to discriminant validity, the work-based SJTs showed lower off-trait correlations with the HEXACO subscales than the school-based SJTs (Openness = .21 vs .28; Integrity = .19 vs .23; Table 11).

Test-Criterion Validity of the School- and Work-Based SJTs

Regarding test-criterion validity, Openness and Integrity showed small but statistically significant correlations with IRB in the school-based version (r = .26 and r = .26, respectively; Table 13). When statistical significance is not considered, both traits showed small positive associations with IRB in both versions, ranging from r = .15 to r = .26. For OCB, correlations were consistently small (r = .13 to r = .18) and non-significant. Compared to Pletzer et al. (2021) meta-analysis, the school-based Integrity-OCB correlation observed here was slightly lower than expected, while the other correlations were similar to what Pletzer and colleagues reported.

Table 13. Pearson Correlation Coefficients Between the School- and Work-Based SJTs and the JPS In-Role Behavior and Organizational Citizenship Behavior Subscales (Study 2)

SJT scale	In-role behavior		Organizational citizenship behavior		Pletzer et al. (2021) meta-analysis
	School-based	Work-based	School-based	Work-based	
Openness	.26*	.15	.17	.17	.20
Integrity	.26*	.19	.13	.18	.21

Note. * = $p < .05$.

Regression analyses were carried out for the IRB and JPS on the school- and work-based relationships. Only the IRB school-based analysis was statistically significant. Regressing OCB on Openness and Integrity accounted for just 9% of the variance ($R^2 = .09$; see Appendix 2 for full model statistics). Both Openness and Integrity showed similar strength of association ($\beta = .22$ and $\beta = .21$, respectively) and almost reached statistical significance.

Level of Confidence Across School- and Work-Based SJTs

Across both SJT contexts, Table 14 shows that students' mean rating of their confidence in their evaluations were high, as at least 76% of respondents described themselves as "quite" or "extremely" confident in their ratings. Openness had the highest confidence in the school setting (79.7%) while Integrity had it on the work setting (84.2%).

Table 14. Mean Confidence Levels of Students in Their Ratings of School and Work-Based SJTs Subscales (Study 2)

Context	SJT scale	Mean confidence in ratings between both SJTs	
		Not at all, slightly, or somewhat confident	Quite or extremely confident
School-based	Openness	20.3%	79.7%
	Integrity	23.6%	76.4%
Work-based	Openness	23.3%	76.7%
	Integrity	15.8%	84.2%

Responses of Students with no Work Experience

Among students with no prior work experience, one-third (34.1%) reported feeling only somewhat confident when judging how they would behave on the job, whereas 41.5% felt quite confident and the rest (24.5%) felt extremely confident.

An analysis of 53 open-ended responses indicated that, despite never having formal employment, most students answered the work-based situations by drawing on similar experiences from school projects, volunteer work, internships, extracurricular activities, and group work. Many relied on their values, personal character, and everyday interactions to guide their responses, imagining themselves in workplace scenarios and answering hypothetically based on how they believe they would act. Others adapted their answers from the school-based

situations provided, treating them as analogous to workplace contexts, or recalling how they had handled comparable situations in other settings.

Discussion

We assume that high school students have familiarity with school-contextualized SJT scenarios, but we cannot assume that they have the same level of familiarity with work-contextualized SJT scenarios. Therefore, we cannot assume that these items will have acceptable reliability and validity evidence for the population of high school test-takers. Our objective with the current study was to compare the reliability and validity of work-based SJTs measuring essential skills vis-à-vis school-based SJTs measuring the same skills, which we considered the gold standard. Taken together, the findings indicate that the work-based SJTs administered to high school students generally demonstrated acceptable reliability and validity evidence to support their use. Across both studies, internal consistency estimates for most work-based scales were in the moderate range, with values often comparable to the school-based versions, and, in several cases, accompanied by adequate convergent and discriminant validity. Criterion-related validity evidence (particularly for Conscientiousness, Agreeableness, and Integrity) aligned with theoretical expectations and prior meta-analytic findings (Pletzer et al., 2021).

However, the results also suggest caution for certain skills. Reliability for work-based Conscientiousness and Integrity was meaningfully lower than their school-based counterparts, and Extraversion demonstrated poor internal consistency across both contexts, which discouraged further validity analyses. In terms of validity, Emotional Stability showed weak and inconsistent criterion relationships, which adds caution in the interpretation of this scale for high school populations. In contrast, Conscientiousness and Agreeableness were the strongest candidates for use in predictive applications, with Openness and Integrity showing more modest but still interpretable evidence.

Student self-reports of confidence add complementary support, as the majority felt “quite” or “extremely” confident in their responses, even when scenarios were work-framed. Importantly, students without prior job experience often mapped workplace scenarios onto familiar non-work contexts (e.g., school projects, extracurriculars), which suggest that lack of employment history did not prevent engagement with the tasks. They also suggest that work-based SJTs are not perceived as unfamiliar or inaccessible, even among students with limited exposure to employment. Their ability to transfer reasoning from school, extracurricular, and personal domains indicates that the work scenarios did not create significant construct-irrelevant variance.

It is also worth noting that lower than expected reliability coefficients may partly be a consequence of restricted variance in the high school population studied. Samples derived from the population of ACT national test-takers are generally high achieving (Walton et al., 2021), and the sampling procedure here relied on self-selection. When samples are relatively homogeneous, such as students close in age, education level, and social context, individual differences in trait expression are compressed, which can underestimate reliability estimates.

This might suggest that some of the observed reliability limitations may be partly because of the narrow range of the population sampled.

While the study supports the feasibility of using work-based SJTs with high school populations, limitations remain, particularly for Emotional Stability, Extraversion, and certain reliability gaps that should be addressed in future development to strengthen the overall validity argument. Given that the SJTs employed here were intentionally generic rather than product-specific, we have confidence that the findings can generalize to a broad array of school- and work-contextualized assessments.

References

ACT. (2024). ACT® WorkKeys® Essential Skills technical manual. ACT, Inc.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345. https://doi.org/10.1080/00223890902935878

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*(1), 83–117.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika, 45*(1), 99–105. https://doi.org/10.1007/BF02293600

Pletzer, J. L., Oostrom, J. K., & de Vries, R. E. (2021). HEXACO personality and organizational citizenship behavior: A domain- and facet-level meta-analysis. *Human Performance, 34*(2), 126–147. https://doi.org/10.1080/08959285.2021.1891072

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI2-XS. *Journal of Research in Personality, 68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Walton, K. E., Radunzel, J., Moore, R., Burrus, J., Anguiano-Carrasco, C., & Murano, D. (2021). Adjectives vs. statements in forced choice and Likert item types: Which is more resistant to impression management in personality assessment? *Journal of Personality Assessment, 103*, 842–853.

Weekley, J. A., & Ployhart, R. E. (Eds.). (2006). *Situational judgment tests: Theory, measurement, and application*. Lawrence Erlbaum Associates.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society: Series B (Methodological), 21*(2), 396–399. https://doi.org/10.1111/j.2517-6161.1959.tb00346.x

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*(3), 601–617. https://doi.org/10.1177/014920639101700305

Appendix

Table A1. Regression Analyses Predicting In-Role Behavior and Organizational Citizenship Behavior from School- and Work-Based SJT Scales (Study 1)

Context	Outcome (DV)	Predictor	B	St. error	β	*t*	*p* value
School-based SJT	In-Role Behavior	Conscientiousness	.401	.103	.445	7.300	<.01
		Agreeableness	.135	.124	.127	1.087	.28
		Emotional Stability	–.027	.085	–.029	–.318	.75
	Organizational Citizenship Behavior	Conscientiousness	.345	.117	.340	2.959	<.01
		Agreeableness	.068	.140	.224	1.908	.06
		Emotional Stability	.058	.096	.056	.928	.54
Work-based SJT	In-Role Behavior	Conscientiousness	.450	.112	.405	4.016	<.01
		Agreeableness	.194	.111	.178	1.742	.08
		Emotional Stability	–.030	.086	–.033	–.345	.73
	Organizational Citizenship Behavior	Conscientiousness	.440	.131	.352	3.348	<.01
		Agreeableness	.142	.131	.116	1.091	.28
		Emotional Stability	.074	.101	.072	.732	.47

Note. Model statistics:
- School-based, IRB: Constant = 3.299, $F(3,89) = 11.540$, $p < .001$, $R^2 = .280$
- School-based, OCB: Constant = 2.142, $F(3,92) = 11.069$, $p < .001$, $R^2 = .272$
- Work-based, IRB: Constant = 2.836, $F(3,89) = 9.872$, $p < .001$, $R^2 = .250$
- Work-based, OCB: Constant = 2.156, $F(3,92) = 6.818$, $p < .001$, $R^2 = .187$

Table A2. Regression Analysis Predicting the In-Role Behavior Subscale from the School-Based SJT Scales (Study 2)

Context	Outcome (DV)	Predictor	B	St. error	β	*t*	*p* value
School-based SJT	In-Role Behavior	Openness	.211	.527	.222	1.968	.05
		Integrity	.178	.107	.213	.188	.06
	Organizational Citizenship Behavior	Openness	.200	.167	.140	1.198	.26
		Integrity	.175	.148	.139	1.186	.24
Work-based SJT	In-Role Behavior	Openness	.122	.115	.123	1.060	.29
		Integrity	.192	.134	.167	.1434	.16
	Organizational Citizenship Behavior	Openness	.223	.174	.149	1.283	.20
		Integrity	.261	.201	.151	1.300	.20

Note. Model statistics:
- School-based, IRB: Constant = 3.925, $F(2,72) = 4.584$, $p = .013$, $R^2 = .088$
- School-based, OCB: Constant = 3.346, $F(2,72) = 1.752$, $p = .151$, $R^2 = .020$
- Work-based, IRB: Constant = 4.188, $F(2,72) = 1.891$, $p = .158$, $R^2 = .024$
- Work-based, OCB: Constant = 2.849, $F(2,72) = 1.064$, $p = .143$, $R^2 = .026$

ABOUT ACT

ACT is transforming college and career readiness pathways so that everyone can discover and fulfill their potential. Grounded in more than 65 years of research, ACT's learning resources, assessments, research, and work-ready credentials are trusted by students, job seekers, educators, schools, government agencies, and employers in the U.S. and around the world to help people achieve their education and career goals at every stage of life. Visit us at www.act.org.