



CritiCall[®]3D

Hire Like Lives Depend On It



Behavioral

Personality Assessments



Operational

Skills-Based Assessments



Situational

Interactive Role-Playing Assessments

DEVELOPMENT & VALIDATION OF CRITICAL 3D

*Predicting Dispatcher Success
and Reducing Turnover Using
Validated Composite Testing*

Biddle Consulting Group, Inc.
606 Sutter Street
Folsom, CA 95630

(916) 294-4250
www.biddle.com





Contents

Overview	3
CritiCall 3D Overview and Its Impact on Hiring and Turnover.....	4
Operational: Skills-Based Assessment	5
Behavioral: Personality Assessments	5
Situational: Interactive Role-Playing Assessments	5
How Should CritiCall 3D Or Its Components Be Used In A Selection Process?	9
What Can Employers Do if They Suspect Adverse Impact May Be Present With Their Use Of CritiCall 3D?	10
Endnotes	12



Overview

CritiCall has been adopted by over 1,400 911/emergency call centers over the last 25+ years and is currently the industry-leading dispatcher assessment in the United States and Canada. CritiCall has also been subjected to rigorous research and validation at several dispatch centers and has demonstrated a consistent and strong connection to job performance in several settings.

As a comprehensive pre-employment assessment, the legacy CritiCall assessment is subject to the limitations that comes with most “deep scan” test batteries: It takes significant time (1–2 hours) to administer to applicants. An additional limitation is that legacy CritiCall focuses only on *operational* skills necessary for the job while leaving other important qualifications untested. While there is no prohibition—legally or otherwise—from measuring only the operational or concrete job skills and abilities, we typically advocate that employers use a well-rounded set of selection procedures in the overall hiring process.

The new **CritiCall 3D** described in this report addresses both of these limitations. CritiCall 3D includes a shorter version of the operational assessment and two additional assessments that combine into an effective battery where the sum is more powerful than the parts. While CritiCall 3D takes about the same amount of time to administer to applicants, this updated battery includes shortened versions of the core, legacy CritiCall tests as well as two additional assessments that measure other important dimensions of applicant qualifications: *situational* skills and *personality* characteristics. This combination of tests has a significantly high correlation to job performance ($r = .634$, explaining about **40%** of job performance) and represents a well-rounded assessment mix that does a great job capturing the “whole applicant” ability set as it relates to the 911 dispatcher position.



CritiCall 3D Overview and Its Impact on Hiring and Turnover

CritiCall 3D is a combined assessment with three equally-weighted tests that target different qualification areas of dispatcher applicants. The **Operational** test is a skill-based assessment that measures the skills and abilities necessary for the first day of the dispatcher job. The **Situational** test (powered by our partner Logi-Serve) measures situational/contextual skills that are necessary for the interpersonal side of being an effective dispatcher. The **Behavioral** test measures personality drivers that have been statistically connected to the 911 dispatcher position in several studies. While CritiCall is based on a Content Validity foundation, all three tests have a proven track record of *criterion-related validity* where the test has demonstrated a statistically significant correlation to job performance to the 911 dispatcher position in more than one study. Each test is described further below.



Operational: Skills-Based Assessment

The Operational test included in CritiCall 3D was derived by extracting shortened versions of our full-length CritiCall assessment. This new derivation is called the **CritiCall OneTest** and it includes shortened versions of the following scales from our full-length CritiCall test: Decision Making, Call Summarization, Data Entry, Memory Recall, Cross Referencing, Position Locating, and Proofreading. Unlike the existing CritiCall test modules, this test assesses multiple domains within one test form. Test takers listen to a simulated call that progressively introduces more information and each of these calls will encompass all the different modules listed above. There are a total of five different scenarios within the OneTest and each scenario follows the same flow and test pattern; only the scenario script/content changes between items.

The full-length CritiCall test has been successfully correlated to job performance in at least two 911 centers, with combined observed and operational validities ($r = .378$ and $.561$,

respectively) and strong, consistent reliability ($r = .91$). The correlation between the non-data entry sections of CritiCall and the OneTest is strong ($r = .50$). This results in a conservative uncorrected validity for the OneTest between $r = .27$ and $r = .31$ and operational validities between $r = .33$ and $r = .37$.¹ The OneTest also has strong reliability ($r = .93$, alpha).



Behavioral: Personality Assessments

The Behavioral component in CritiCall 3D was developed by ioPredict, a test development and validation partner of Biddle Consulting Group, Inc. The content of the Behavioral component is proprietary, but it can be generally stated that it is based on the conventional 5-factor personality model using likert-type, poly-weighted items. As part of developing the updated CritiCall 3D assessment a refresh was conducted on this set of personality items that improved the validity and decreased the number of items from 60 to 40.² The final scale was developed using a split-half/hold-out type study that resulted in the final 40-item scale demonstrating strong uncorrected validity ($r = .321$) and reliability ($r = .732$).



Situational: Interactive Role-Playing Assessments

The Situational component is powered by Biddle Consulting Group's assessment partner, Logi-Serve. This assessment provides various interpersonal scenarios and requests the applicant to select the most appropriate response to each. It includes an advanced branching part of the assessments where applicants are also asked to provide the likely outcomes of various behavior choices and to forecast various outcomes. The assessment also includes a personality-based component.

The Logi-Serve assessment has been previously validated in 911 dispatch centers ($r = .415$, $n = 116$) and an operational validity of $r = .549$ (corrected based on moderate, indirect range restriction ($u = .70$) and criterion unreliability assuming a criterion reliability of $.70$). The test also has strong reliability ($r = .82$; test-retest).

Combined Assessment

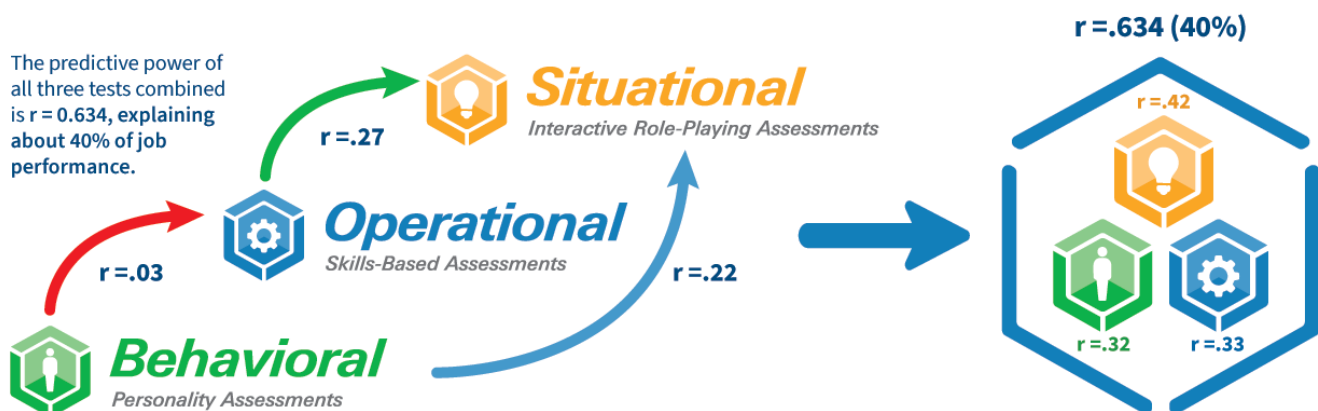
All three components (Operational, Behavioral, and Situational) were equally weighted and combined to create the CritiCall 3D test. A significantly high correlation emerges when these tests are combined for two reasons. First, each test is significantly correlated to the job independently. Second, there is limited overlap between the tests, as revealed in the Table below.

Test		Operational	Behavioral	Situational
Operational	Pearson Correlation	1	0.03	.219**
	Sig. (2-tailed)		0.608	0
	N	301	301	301
Behavioral	Pearson Correlation	0.03	1	.219**
	Sig. (2-tailed)	0.608		0
	N	301	623	301
Situational	Pearson Correlation	.219**	.219**	1
	Sig. (2-tailed)	0	0	
	N	301	301	301

** . Correlation is significant at the 0.01 level (2-tailed).

These intercorrelations are displayed graphically in the Figure below.

DISPATCHER JOB PERFORMANCE



When all three tests are combined into a composite, a remarkably strong validity coefficient of $r = .634$ emerges, explaining about 40% of job performance variance³ (determined by squaring the .634 coefficient). In other words, about 40% of job performance is captured and explained by the combined set of Assessments. As displayed in the Table below, correlations that exceed .35 are classified as “very beneficial” by the U.S. Department of Labor (U.S. DOL, 2000).

Table - U.S. Department of Labor Guidelines for Interpreting Validity Coefficients.

U.S. DEPARTMENT OF LABOR GUIDELINES FOR INTERPRETING VALIDITY COEFFICIENTS		
Coefficient Value	Interpretation	Percentage of Job Performance Explained
Above .35	Very beneficial	12.3%+
.21 - .35	Likely to be useful	4.4% - 12.3%
.11 - .20	Depends on circumstances	1.2% - 4%
Below .11	Unlikely to be useful	<1.2%

Source: U.S. Department of Labor (2000), Testing and Assessment: An Employer's Guide to Good Practices (p. 3-10)

When weighing the credibility of validity studies in legal settings, some courts have ruled that correlations that exceed .30 surpass the threshold required for justifying the use of tests on a strict, rank-ordered basis.⁴ Squaring a .30 correlation results in 9% explanatory power for job performance. Squaring the .634 correlation of the CritiCall 3D test results in 40%, about 4.4X stronger than some courts have required for demonstrating a sufficiently strong correlation for ranking on a selection procure? Because of the strength of this correlation, 911 dispatch centers that use CritiCall 3D are likely to improve the effectiveness of their hiring program¹ from 50% to 84% (when an assumption is made that 50% of the applicants who apply for employment are qualified and a 20% cutoff is used for CritiCall 3D, or passing 80% of the applicants).

There are two advantages for using CritiCall 3D: (1) **Screening out** the dispatcher candidates who are most likely to turnover for job-performance related reasons, and (2) **Selecting in** the dispatcher candidates who will most likely perform well on the job. CritiCall 3D is able to strategically and simultaneously accomplish these two goals because of the exceptionally high correlation ($r = .634$) with job performance that results when the three tests that make up CritiCall 3D are combined into a single, powerful assessment tool.

The best way to explain this benefit is by providing an example (see Appendix). These examples were computed using the **CritiCall 3D Job Performance and Turnover Calculator**, available online.

¹ When compared to using no testing program, or a selection process with invalid tests.

Consider an Emergency Dispatch Center that has **20** dispatchers begins using CritiCall 3D for hiring incoming candidates with the following factors and assumptions:

- 50% “**Qualification Rate**” will be assumed for the incoming emergency dispatchers, meaning about one-half of the incoming applicants possess the “baseline” levels of qualifications needed for successful job performance. This can also be considered the rate at which unscreened candidates would successfully pass training, academy, probation, and perform on the job.
- 50% **Passing Rate** will be used for CritiCall 3D (resulting with about half of the applicants passing onto the next step in the hiring process).
- \$54,340 as the Annual Emergency Dispatcher **Salary** (w/o benefits) (see Appendix).
- 29% annual **turnover** rate.⁵
- 6 year average **tenure**.

Using the assumptions above, this Center will increase the 50% training/probationary **success** rate to 72% and the training/probation **failure** rate will decrease from 50% to 28%. Their average turnover will decrease from 29% to 22.8%, and their annual separations will decline from 5.8/year to 4.6/year. Their annual unplanned turnover cost will decrease from \$786,888 to \$617,959, resulting in a 5-year savings of \$844,647, a 21% decrease in costs associated with unplanned turnover.

While the above scenario assumes the agency used a 50% cutoff score, the benefits increase substantially when a stronger cutoff is used, such as hiring only the applicants who score in the top 20%. In this situation, the agency will increase the 50% training/probationary **success** rate to 86% and the training/probation **failure** rate will decrease from 50% to 14%. Their average turnover will decrease from 29% to 18.6%, and their annual separations will decline from 5.8/year to 3.7/year. Their annual unplanned turnover cost will decrease from \$786,888 to \$503,713, resulting in a 5-year savings of \$1,415,873, a 36% decrease in costs associated with unplanned turnover.



How Should CritiCall 3D Or Its Components Be Used In A Selection Process?

Dispatch centers that use CritiCall 3D for positions that are highly similar to the “mainstream” (very typical) 911 emergency dispatcher position that was the focus of the studies upon which it was developed can use scores from the combined CritiCall 3D assessment in three ways:

- Strict rank ordering,
- Top-down banding, or
- Pass/fail cutoff.

These three options will likely result in the highest-to-lowest adverse impact ratios as well as the highest-to-lowest collective dispatch center-wide job performance levels. Perhaps the most balanced approach is top-down **banding**, which groups applicants in terms of their qualification levels while considering the non-true score variability of their scores. In other words, banding uses confidence interval brackets to group applicants who have similar abilities while acknowledging that tests imperfectly measure the targeted skills and abilities, knowing that if all applicants repeated the tests their scores would likely shuffle within predictable ranges.

Dispatch centers that wish to use any of the three components in CritiCall 3D individually are encouraged to complete our Validity Transportability process to evaluate whether ranking is appropriate.

While these recommendations are “global” in nature, we recommend contacting Biddle Consulting Group for situations where the target position is dissimilar from the “mainstream” dispatcher position or adverse impact is present in the hiring process.



What Can Employers Do If They Suspect Adverse Impact May Be Present With Their Use Of CitiCall 3D?

We recommend that all employers conduct a “validation transportability” study with our assistance. This is especially true of employers that use CitiCall 3D in a way that exhibits adverse impact. A ground-up validation study (using content or criterion-related validity for the Operational scale) and criterion-related validation studies for the Behavioral and Situational requires larger sample sizes and is not always possible because of range restriction (especially where CitiCall 3D or a similar test was used for screening applicants). For this reason, we suggest opting for the ease of transporting the criterion validity through our transportability study template.

The transportability study can be conducted by using surveys made available by our organization and requires a few hours of survey work by a panel of 12 subject matter experts (SMEs) (typically incumbents with at least one year of experience and a few supervisors). This survey will enable us to evaluate whether the ratings from your SMEs sufficiently address the validation transportability requirements of Section 7B of the *Uniform Guidelines on Employee Selection Procedures*.

Appendix

Typical Turnover Cost per 911 Dispatcher		
Cost	Source	Description
\$ 5,000	Recruitment and onboarding	Advertising, background checks, pre-employment testing, drug testing, HR and supervisor time.
\$ 53,253	Dispatcher in training/job for training/probation period	Payroll and benefits for dispatcher to complete training and begin work within training/probation months. Includes academy or classroom training, CTO (Communications Training Officer) hours, Certifications (NCIC, CPR, EMD, etc.), and trainee wages during non-productive time. 1.4X burden cost also added to base salary.
\$ 23,964	Dispatcher training	CTO taken off full dispatch duties, supervisor oversight time, reduced call-taking capacity. Includes 30% fraction of Dispatcher Trainer cost (to account for multiple trainees in training concurrently). Cost was determined by training/probation months of Center Dispatcher Trainers, who typically earn 50% more than entry-level dispatchers (1.4X burden cost also added to base salary).
\$ 53,253	Backfill	Pay for dispatcher to fill empty position of trainer. Other costs may include OT for remaining dispatchers; increased fatigue leading to higher error and sick-leave rates.
\$ 200	Training supplies	Books, binders, etc.
\$ 135,670	Total	
	Override Amount (Leave Blank to Use Default)	
Training Investment Hours:		1271
Percentage of Year:		70%

Endnotes

¹ While the mechanically estimated true score overlap between CritiCall and the OneTest ($k \approx 0.542$) results in an estimated OneTest validity of $r = .21$, because the OneTest is essentially a derivative of the core tests in CritiCall, the OneTest is likely to retain 70–80% of the criterion-relevant variance from CritiCall. Given this reasonable assumption, the likely uncorrected validity coefficient for the OneTest was provided in the report. Correcting for direct range restriction (because CritiCall was based on an incumbent sample, and assuming a typical 18% Selection Ratio, the validity is between $r = .33$ and $r = .37$, so the lower value was used in this study when estimating the overall validity for CritiCall 3D. This value is likely low because the CritiCall validity was $r = .378$ and correcting this value for Direct Range Restriction of just 50% provided a validity coefficient of $r = .561$. However, because the OneTest is a distilled version of this longer test, the lower value was used, which is also likely to be observed in an actual study.

² The previous set of personality items included 60 scored items and 20 experimental items and was based on a criterion-related study with significant results in 2021. The current split-half/hold-out study included 623 subjects and also 20,837 job applicants (for variability research with the goal of having 25%, 50%, and 25% in the three groups). The new test has 40 items and 10 experimental items. Each of the 4 subscales have 8 unique items and there are 8 items not assigned to any subscale.

³ The resulting validity coefficient from combining these three assessments is $r = 0.531$ (uncorrected composite validity). The criterion-corrected composite validity (assuming $ryy = .70$) is: $r = 0.634$, explaining about 40% of job performance. If the CritiCall long-form test is included in the three-test battery, this likely increases to (uncorrected composite validity) = 0.646 with a criterion-corrected composite validity (assuming $ryy = .70$) of $r = 0.772$, explaining about 60% of job performance.

⁴ Biddle, D. A. (2006). Adverse impact and test validation: A practitioner's guide to valid and defensible employment testing. Burlington, VT: Ashgate.

⁵ An annual turnover rate of 29% is used as a default value (see APCO Project RETAINS (Responsive Efforts To Address Integral Needs in Staffing: <https://www.apcointl.org/services/staffing-retention/project-retains/#study>).