Measuring Critical Thinking, Learning Agility, and Distance Traveled

Kylie Goodell King, Ph.D.

CLARA™

Abstract

This report describes the development, testing, review, and validation of a new instrument meant to assess job candidate's ability in the following three areas: critical thinking, learning agility, and distance traveled. Results indicate that the instrument demonstrates validity in some areas and further exploration is needed in others.

Critical Thinking, Learning Agility, and Distance Traveled Validation Report

The workplace of today and the future requires successful employees to possess different skills than those of the past (Mindell & Reynolds, 2023). The ability to think critically, practice learning agility, and leverage the benefits of distance traveled will enable individuals to make a positive impact in dynamic organizations as technologies rapidly increase the pace of work and rate of change (Carnevale & Smith, 2013; DeMeuse, 2017; Grant, 2023; Mindell & Reynolds).

This report describes the development, testing, review, and validation of an instrument that measures the critical thinking ability, learning agility, and appreciation for distance traveled of prospective employees. The ultimate objectives of this assessment are to evaluate the abilities of job candidates to successfully contribute in workplaces that require skill in these three construct areas and to provide information to individuals and organizations about areas of strength and opportunities for improvement.

Introduction

Critical thinking, learning agility, and distance traveled are constructs that influence the ability for an individual to succeed in dynamic work environments (Carnevale & Smith, 2013; DeMeuse, 2017; Dewson, Eccles, Tackey & Jackson, 2000). These three constructs measure employee attributes that may be used to predict future work performance and organizational fit.

Critical thinking, defined here as the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action, has been a popular idea in organizational parlance and practice for over half a century (e.g. Smith, Knudsvig & Walter, 1960). Learning agility, which focuses on human behavior, high-level cognitive processing, and the selective transference of lessons learned in one setting and applying them to a uniquely different one, developed as a potentially important concept more recently, in the early 2000s (e.g. Milani, Setti & Piergiorgio, 2021). Distance traveled, which is defined here as a measure of an individual's accomplishments and capacities that considers one's starting point and trajectory of progress,
has emerged over the past decade as a construct that might influence workplace success in certain environments (Ray & Brown, 2015).

Purpose of the Report

The purpose of this report is to describe the development, testing, review, and validation of an instrument that measures critical thinking, learning agility, and distance traveled in prospective employees. This assessment is intended to be used to evaluate job candidates and provide prospective employees with information about their candidacy that they can use in future job searches.

Overview of the Assessment

The proposed instrument is a long-form version of the assessment which included 37 items; 12 items to

measure each of six facets of distance traveled (two items per facet; facets include grit, resourcefulness, self-efficacy, proactivity, resilience, and antifragility), 10 items to measure five facets of learning agility (two items per facet; facets include cognitive agility, people agility, change agility, results agility, and self-awareness), and 15 items to measure five facets of critical thinking (three items per facet; facets include evaluating, problem solving, inferring, analysis, and asking questions).

The assessment also includes three self-assessment items related to each of the three constructs and a number of (optional) items that provided demographic information about participants. A copy of the assessment is included in Appendix A.

Definition of Validity

Validity refers to the extent to which a proposed instrument accurately measures the constructs it is intended to assess. In this case, validity would illustrate that the instrument accurately measures critical thinking, learning agility, and distance traveled and would reflect the degree to which evidence and theory support the interpretation of assessment results as a contributing factor to the evaluation of job candidates. A comprehensive review of validity ensures that the results of this assessment are meaningful and appropriate for the context of candidate evaluation.

Assessment Development

Item Generation

The item generation process was different for each of the three evaluated constructs based on prior development of measurement instruments and construct validation. Critical thinking and learning agility items were constructed to follow previously validated items and instruments. For example, Norris (1989) describes the validation of a critical thinking instrument and DeRue and colleagues describe a process of theoretically grounding a conceptualization of learning agility (2012). Critical thinking items were developed to be in line with the Watson-Glaser Critical Thinking Appraisal, tapping facets of critical thinking including problem solving, analysis, evaluating, asking questions, and inferring (Hassan & Madhum, 2007; Paul & Elder, 2019). Learning agility was conceptualized based on the work of Lombardi and Eichinger (2000) and the later viaEDGE assessment, as summarized by De Meuse (2017).

The development of distance traveled items required a more creative and generative approach which was informed by an extensive literature review, including works from diverse fields such as higher education (e.g. Bastedo, Bowman, Glasener & Kelly, 2018; Craig, 2017; Stephens, Fryberg, Markus, Johnson & Covarrubias, 2012;), human resources and organizational behavior (e.g. Dewson, Eccles, Tackey & Jackson, 2000; Grant, 2023; Taras, Gunkel, Assouad, Tavoletti, Kraemer, Jimenez, Svirini, Lei & Shah, 2021), organizational psychology (e.g. Duckworth, Eichstaedt & Ungar, 2015; Hough, 1984; McNeish, Dumas, Dong & Duellberg, 2023; Ray & Brown, 2015; Williams & Gilovich, 2012), medical education (Ellsworth, Solano, Evans, Bidwell, Byrnes & Sandhu, 2024; Epstein, Konuthula, Meyer, Whipple, Bowe, Bly & Abuzeid, 2022; Saguil & Kellerman, 2014; Smedley, Butler & Bristow, 2004), and even the classic autobiographical work of Booker T. Washington (1907).

Items were reviewed by and input was provided from a team of experts with backgrounds in data science, entrepreneurship, human resources, software development, marketing, psychology, and machine learning.

Item Scoring

Items scores were reviewed by a team of experts with relevant expertise. Most critical thinking items were scored as either correct (one point awarded per item) or incorrect (zero points awarded per item) with the exception of "Inferring" items (see Appendix A) which awarded points for selecting an incorrect, though reasonable, response. Learning agility items were all scored based on a scoring matrix which awarded one point for the most correct response and up to 0.5 points for responses that were also deemed reasonable by a team of experts. Distance traveled items were also scored based on a scoring matrix which awarded one point for the most correct response and up to 0.75 points for other reasonable responses. This scoring scheme reflects the level of ambiguity that is inherent in the measurement of distance traveled, a relatively new construct, and of learning agility, a construct which has been measured but is still undergoing continued development and consideration.

Pilot Testing and Final Item Selection

The content experts reviewed the initial instrument and provided feedback about the instrument length, formatting, and composition that was used to modify the original assessment. A major change that was suggested through this testing was to ensure that the reading level of the assessment was appropriate for potential assessment participants. This feedback was incorporated into the final instrument, which is written at a fifth grade reading level.

Data Gathering

Participants

Participants were identified, completed the assessment, and were paid through the Positly platform (Positly, 2024). Participants were selected using the "Microtasker" designation and were limited to taking and receiving compensation for one assessment. Participant quality was ensured by blocking suspicious IP addresses, requiring attentive participants, requiring high approval rate and experience, having a minimum Positly platform activity approval rate, and having completed a number of previous Positly tasks.

Participants were limited to the United States. Ages ranged from 20 to 79 years old. The median age of participants was 42 years. 48% of participants identified as female and 52% as male. The median participant income was $72,500, slightly above the national average income of $69,800 (SoFi, 2024).

Of approximately 8,000 eligible Positly participants, 385 completed the assessment. Upon an initial review of the data from these participants, 12 were excluded resulting in 373 completed assessments. Based on an estimated US workforce of 160 million individuals (Statista, 2024), this sample provides a level of confidence of 95% and a margin of error of +/- 5%.

Instrument

The assessment instrument includes 37 items; 15 items meant to provide information about five facets of

critical thinking, 10 items meant to provide information about five facets of learning agility, and 12 items meant to provide information about six facets of distance traveled. The instrument was developed using the GuidedTrack platform (GuidedTrack, 2024). The instrument also includes three self-assessment items related to each of the three constructs and a number of (optional) items that provided demographic information about participants, including age, level of education, income, veteran status, student status, and whether the participant is an early-career employee, re-entering the workforce after an extended absence, or considering a career change. The assessment instrument is provided in Appendix A.

Evaluation of Validity

Multiple forms of validity were evaluated in the assessment of the proposed instrument, including Content validity, construct validity, criterion validity and face validity.

Content Validity

In order to evaluate the assessment for relevance and representativeness of the three constructs(critical thinking, learning agility, and distance traveled), a number of theoretical models and frameworks were reviewed and incorporated into construct definitions, facet descriptions, and item development (e.g. De Meuse, 2017; Eichinger & Lombardo, 2004; Ellsworth et al., 2024). This process ensured that the resulting items are reflective of the most widely used and agreed-upon conceptualizations of critical thinking, learning agility, and distance traveled.

Construct Validity

In order to establish and evaluate the internal structure of the assessment, a confirmatory factor analysis was undertaken using the R software platform (R Core Team, 2024). Results of this analysis, including the model illustration provided in Appendix B, indicate that a three factor model appropriately fits the gathered data. The analyzed sample includes the 373 Positly participants described above. A number of assumptions were checked and confirmed including the adequacy of the sample size, linearity, and multicollinearity.

Principal component analysis was used to extract factors followed by an oblimin rotation. The results of a scree plot / parallel analysis, indicate that a three factor model was appropriate for the data. Using a minimum factor loading of 0.20, Table 1 below illustrates the interpretation of factors and item loadings. Table 2 provides the correlations between factors, and table three provides a summary of the variance explained by each of the three factors. Appendix B includes the scree plot and full factor loadings.

Table 1

Interpretation of Factors and Factor Loadings

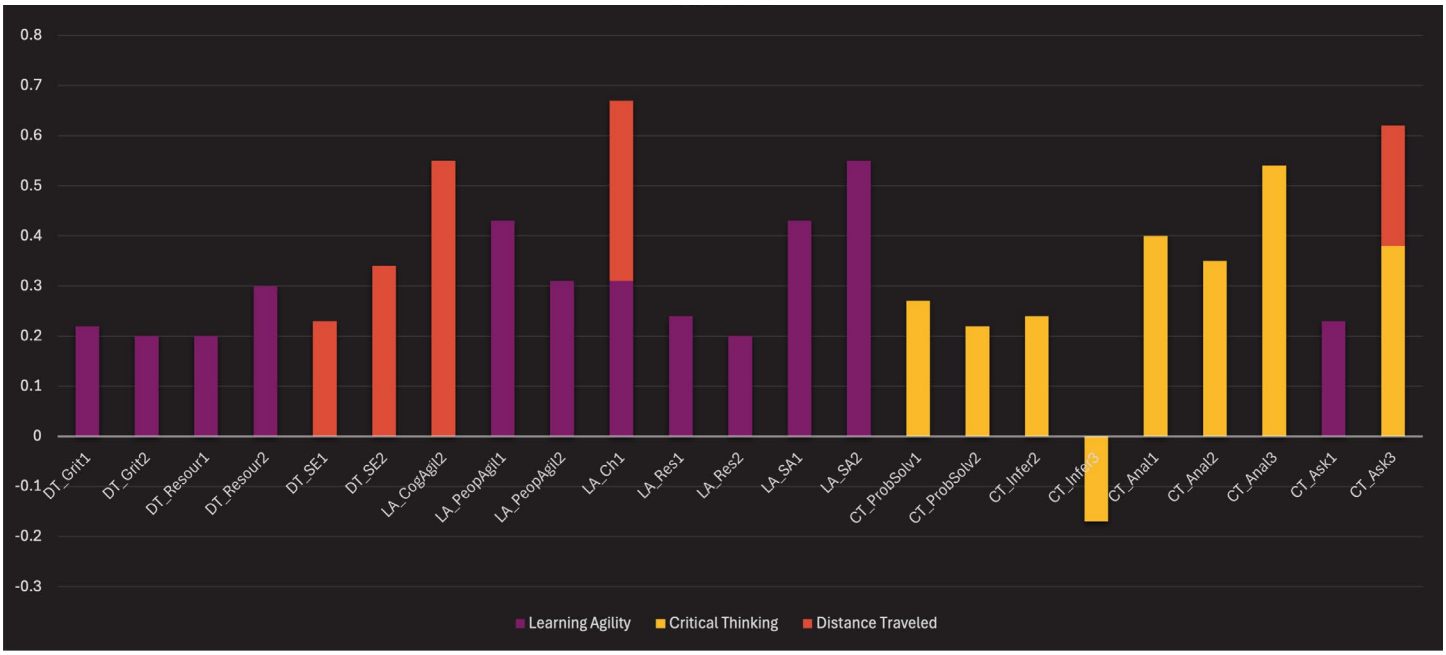


Table 2

Factor Correlations

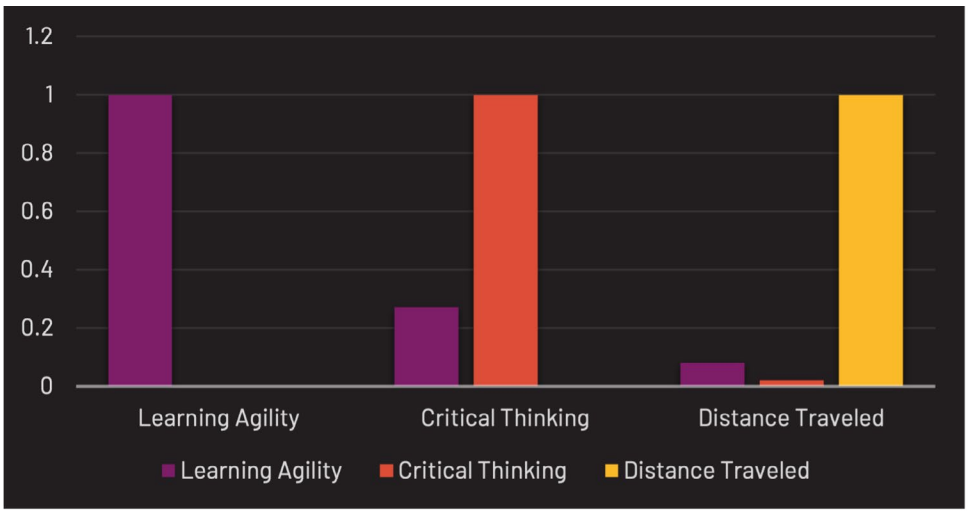
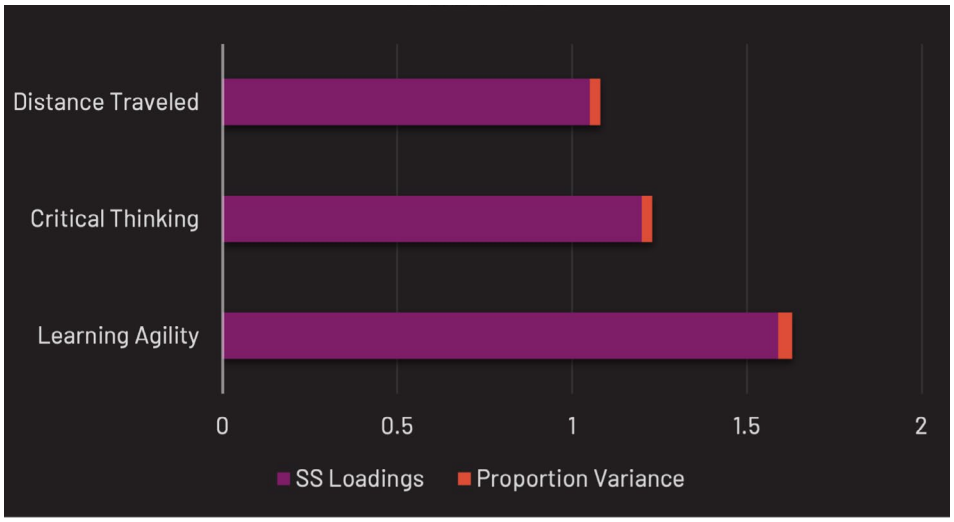


Table 3

Variance Explained



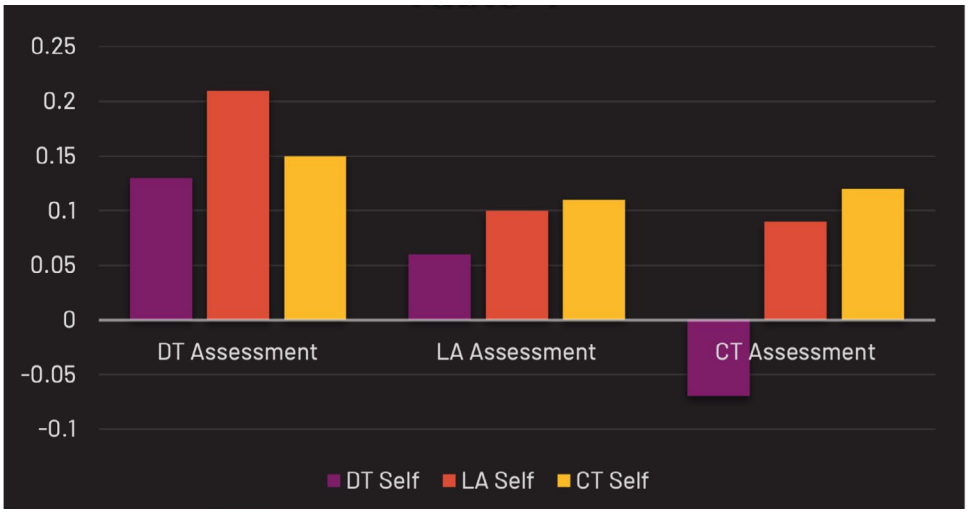
A number of fit indices were evaluated including the Tucker-Lewis Index (TLI) and root mean square residual (RMSR). The RMSR is equal to 0.05, indicating good fit; however, the TLI is equal to 0.81, which is below the value of 0.90 which is typically used to suggest reasonable fit. Based on these loadings, along with the other information described above, some evidence for construct validity is provided; however, further analyses should be undertaken after adhering to some of the limitations described later in this report.

Convergent, Discriminant, Criterion, and Concurrent Validity

Convergent validity was evaluated by examining how assessment ratings compared to self-ratings. Results of this analysis are detailed below in Table 4. There were very weak correlations between each of the measured constructs and self evaluations. For critical thinking, correlations were higher between the critical thinking self-evaluation and the items meant to measure critical thinking. For learning agility and distance traveled, self-evaluations were more similar to other constructs.

Table 4

Correlations Between Self-Ratings and Assessment Ratings



Although this does not provide evidence of convergent validity, it is unsurprising given the lack of precision with which individuals often evaluate their own traits (e.g. Gao, Saiedur Rahaman, Shao & Salim, 2021; Kormos & Gifford, 2014; Paunonenn & O'Neill, 2010). Recommendations for further evaluations of convergent and discriminant validity, as well as potential evaluations of criterion-related, or predictive validity, are discussed later in the recommendations section of this report. Evaluations of concurrent validity are also discussed.

Face Validity

Leveraging feedback from the team of expert reviewers previously described, it was determined that the instrument holds face validity, that it appears to measure what it intends to measure. Although not a rigorous form of validity, this feedback provides useful information about how the assessment may be perceived by future participants.

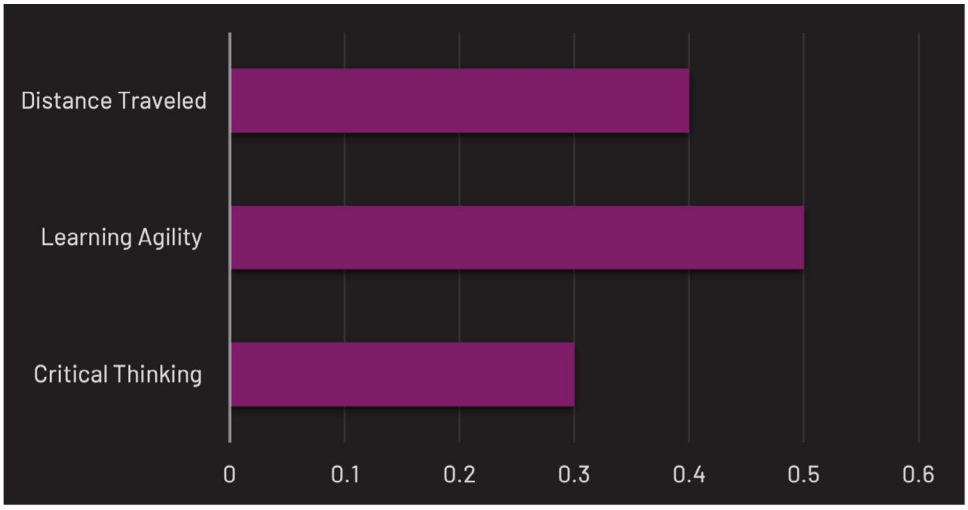
Evaluation of Reliability

Internal Consistency

Cronbach's alpha provides an informative measure of internal consistency (Cronbach, 1951). Expressed as a number between 0 and 1, this measure evaluates the extent to which all items in a subset measure the same construct. Values for Cronbach's alpha for the three evaluated constructs are provided below in Table 5.

Table 5

Cronbach's Alpha



Although these values are all below the threshold for typically acceptable values of 0.6, there are reasons for which this boundary is not achieved. There is an inherent tradeoff between internal consistency and the time it takes to complete the assessment. In this case, brevity was favored over potential issues of reliability. Furthermore, the three evaluated constructs are not theorized to be unidimensional; therefore, it is somewhat expected that internal consistency would be rather low. For example, it is likely that, although conceptually different, the facets of cognitive agility and people agility, both subsets of learning agility, are related to each other.

Test-Retest Reliability

Given the timeframe and purpose of this instrument, this exercise did not attempt to evaluate the reliability of this instrument by retesting participants; however, this could be a useful exercise to complete in the future.

Implications for Use

Appropriate Uses of the Assessment

This assessment is meant to be used in the evaluation of candidates for employment. Candidates are assumed to be at least 18 years of age and applying for jobs in the United States. The assessment was crafted to be appropriate for those individuals with a fifth-grade or higher reading level. Single or multiple constructs may be evaluated given the type of role for which candidates are being evaluated and preferences of the hiring organization.

The assessment may also be used to provide job candidates with information about their capabilities in these three construct areas. It is theorized that all three constructs, critical thinking, learning agility, and

distance traveled, may be improved over time; therefore, this assessment might be useful for candidates to understand how they perform in each of these three areas so that improvements might be pursued.

Interpretation of Scores

Employers and candidates will be provided with an interpretation of scores following the assessment. For critical thinking and learning agility, they will be given a summary of strengths (high scores) and opportunities for growth (low scores) in considering the various facets of each of the two constructs. As distance traveled is a construct that is still undergoing theoretical development, candidates and employers will be provided with a summary of facets of this construct for which they had relatively high scores.

Potential Misuses

This assessment should not be used singularly to qualify or disqualify candidates for positions. It is meant to provide one set of measurements to indicate candidate potential in terms of critical thinking, learning agility, and distance traveled.

Discussion and Conclusion

Summary of Validity Evidence

The validation process described above provides strong support for certain forms of validity including content validity and face validity and some evidence of construct validity and internal consistency. Convergent, discriminant, criterion, and concurrent validity as well as test-retest reliability should be further evaluated in future analyses.

Limitations of the Validation Process

Although the instrument demonstrates some evidence of validity, there are a number of limitations that should be considered. The first limitation comes with collecting data through a third-party platform. Although the Positly participants were, as described, selected for the quality of their responses, these participants did not have a high level of motivation in completing this assessment; therefore, it is likely that responses were not as thoughtfully provided as they would be in a pre-employment setting. It would be useful to collect data from participants who have a greater sense of motivation in providing more contemplative responses.

Recommendations for Future Research

As previously described, there is a limitation in using self-reports to establish convergent validity. Future evaluations of this assessment might consider including ratings from others (e.g. supervisors, peers) in order to better establish the convergent validity of the instrument. These measures could also provide evidence of criterion-related validity if they include assessments of the contributions of these individuals in their places of work. Concurrent validity could also be evaluated if this assessment is to be used in conjunction with items from existing instruments that measure learning agility and critical thinking; although pre existing measures of distance traveled are not well established. A longer instrument could also be created and evaluated to see if the supposed tradeoff between brevity and internal consistency exists.

Next Steps

We are committed to continuously testing and refining our tool over time to ensure its accuracy and relevance in diverse hiring environments. Currently we are running validation tests with a small number of pilot talent acquisition teams. In addition, we are actively seeking further academic validation from experts in the field to enhance the credibility of our approach. As part of this ongoing process, we will also be conducting a deeper analysis of the data to identify trends and insights across specific demographic groups. Updates will be made to this report as additional information becomes available.

References

Bastedo, M. N., Bowman, N. A., Glasener, K. M., & Kelly, J. L. (2018). What are we talking about when we talk about holistic review? Selective college admissions and its effects on low-SES students. The Journal of Higher Education, 89(5), 782-805. : https://www.jstor.org/stable/10.2307/26772134

Carnevale, A. P., & Smith, N. (2013). Workplace basics: The skills employees need and employers want. *Human Resource Development International*, *16*(5), 491-501.

Craig, R. (2017). A "distance-traveled" model for college admissions. Inside Higher Ed.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

De Meuse, K. (2017). Learning agility: Its evolution as a psychological construct and its empirical relationship to leader success. , *69*(4). 267-295. http://dx.doi.org/10.1037/cpb0000100

DeRue, D. S., Ashford, S. J., & Myers, C. G. (2012). Learning agility: In search of conceptual clarity and theoretical grounding. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 5*, 258 –279. http://dx.doi.org/10.1111/j.1754-9434.2012.01444.x

Dewson, S., Eccles, J., Tackey, N. D., & Jackson, A. (2000). Guide to measuring soft outcomes and distance travelled. Institute for Employment Studies, Brighton, UK.

Duckworth, A. L., Eichstaedt, J. C., & Ungar, L. H. (2015). The mechanics of human achievement. Social and Personality Psychology Compass, 9(7), 359-369. doi: 10.1111/spc3.12178

Ellsworth, B. L., Solano, Q. P., Evans, J., Bidwell, S. S., Byrnes, M., & Sandhu, G. (2024). Medical students' perception of their 'distance traveled' in medical school applications. Medical Education. https://doi.org/10.1111/medu.15167

Epstein S, Konuthula N, Meyer TK, Whipple ME, Bowe SN, Bly RA & Abuzeid, WM. (2022). Implementing a "distance traveled" question to improve resident diversity: Process and feasibility. *OTO Open, 6*(3). doi: 10.1177/2473974X221113847.

Gao, N., Saiedur Rahaman, M., Shao, W., & Salim, F. D. (2021). Investigating the reliability of self-report data in the wild: The quest for ground truth. In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers.

Grant, A. (2023). Hidden Potential: The Science of Achieving Greater Things. Penguin.

GuidedTrack. (2024). GuidedTrack: Platform for creating interactive programs. GuidedTrack. https://www.guidedtrack.com.

Hassan, K. E., & Madhum, G. (2007). Validating the Watson Glaser critical thinking appraisal. *Higher Education, 54*, 361-383. https://doi.org/10.1007/s10734-006-9002-z

Henderson, M. C., Fancher, T. L., & Murin, S. (2023). Holistic admissions at UC Davis—journey toward equity. JAMA. 2023;330(11):1037-1038. doi:10.1001/jama.2023.15872

Hough, L. M. (1984). Development and evaluation of the" accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69*(1), 135-146. https://doi.org/10.1037/0021-9010.69.1.135

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology, 40*, 359-371. https://doi.org/10.1016/j.jenvp.2014.09.003

Lombardo, M. M., & Eichinger, R. W. (2000). High potentials as high learners. *Human Resource Management, 39*(4), 321–330. https://doi.org/10.1002/1099-050X(200024)39:4<321::AID-HRM4>3.0.CO;2-1

McNeish, D., Dumas, D., Dong, Y., & Duellberg, D. (2023). Promoting inclusive recruiting and selection into military training schools: Admission waivers versus retesting. *Journal of Applied Psychology, 109*(3), 415–436. https://doi.org/10.1037/apl0001147

Milani, Setti & Argentero. (2021). Learning agility and talent management: A systematic review and future prospects. *Consulting Psychology Journal: Practice and Research, 73*(4). 349-371. 10.1037/cpb0000209.

Mindell, D. A., & Reynolds, E. (2023). The work of the future: Building better jobs in an age of intelligent machines. MIT Press.

Norris, S. P. (1989). Can we test validly for critical thinking?. *Educational researcher, 18*(9), 21-26. tps://doi.org/10.3102/0013189X01800902

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools.* Rowman & Littlefield.

Paunonen, S. V., & O'Neill, T. A. (2010). Self-reports, peer ratings and construct validity. *European Journal of Personality: Published for the European Association of Personality Psychology, 24*(3), 189-206. https://doi.org/10.1002/per.751

Positly. (2024). Positly: Online platform for participant recruitment. Positly. https://www.positly.com

R Core Team. (2024). R: *A language and environment for statistical computing* (Version 4.3.0) [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ray, R. & Brown, J. (2015). Reassessing student potential for medical school success: Distance traveled, grit, and hardiness, *Military Medicine,180*(4), 138–141. https://doi.org/10.7205/MILMED-D-14-00578

Saguil, A., & Kellermann, A. L. (2014). The community college pathway to medical school: A road less traveled. *Academic Medicine, 89*(12), 1589-1592. DOI: 10.1097/ACM.0000000000000439

Smedley, B. D., Butler, A. S., & Bristow, L. R. (2004). Increasing diversity in the health professions: a look at best practices in admissions. In In the Nation's Compelling Interest: Ensuring Diversity in the Health-Care Workforce. National Academies Press (US).

Smith, B. O., Knudsvig, G., & Walter, T. (1960). Critical thinking. Recent research developments and their implications for teacher education. 13th Yearbook, American Association of Colleges for Teacher Education. Washington, DC: The Association.

Smith, B. A., & Watkins, K. E. (2024). Measuring learning agility: a review and critique of learning agility measures. *Personnel Review, 53*(3), 704-720. https://doi.org/10.1108/PR-10-2023-0886

Statista. (2024). *U.S. civilian labor force seasonally adjusted 2022-2024.* Statista. https://www.statista.com/statistics/193953/seasonally-adjusted-monthly-civilian-labor-force-in-the-us/

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of

first-generation college students. *Journal of Personality and Social Psychology, 102*(6), 1178-1197. DOI: 10.1037/a0027143

SoFi. (2024). Average salary in the U.S. 2024. SoFi. https://www.sofi.com/learn/content/average-salary-in-us/

Taras, V., Gunkel, M., Assouad, A., Tavoletti, E., Kraemer, J., Jiménez, A., Svirina, A., Lei, W.S. and Shah, G. (2021) 'The predictive power of university pedigree on the graduate's performance in global virtual teams', *European Journal of International Management, 16*(4), 555–584. https://doi.org/10.1504/EJIM.2021.118582

Washington, B. T. (1907). Up from Slavery: An Autobiography. Doubleday.

Williams, E. F., & Gilovich, T. (2012). The better-than-my-average effect: The relative impact of peak and average performances in assessments of the self and others. Journal of Experimental Social Psychology, 48(2), 556-561.