*Acta Educationis Generalis*
*Volume 16, 2026, Issue 2*

DOI: 10.2478/atd-2026-0010

# Academic Performance and Success: Rethinking Educational Metrics

***Pascal Roubides****

Received: December 8, 2025; received in revised form: February 28, 2026; accepted: March 3, 2026

**Abstract:**
***Introduction:*** This article discusses results emanating from a study examining the validity of current measures of academic success in higher education.
***Methods:*** The study uses a quasi-experimental quantitative design to compare outcomes of traditional online College Algebra courses and the same course using a hybrid competency-focused model with a polyparadigm delivery approach.
***Results:*** Preliminary findings showed differences in the proportions of successful course completions based on how "success" was interpreted.
***Discussion:*** These findings imply that conventional performance-based metrics may not adequately represent authentic learning or student academic growth.
***Limitations:*** Limitations include substantial differences in sample sizes being compared and confounding variables related to the course delivery design and implementation, even the actual metrics being measured.
***Conclusions:*** A broader reconceptualization of academic success is necessary, balancing quantitative and qualitative measures to capture the complexity of learning in digital learning environments.

**Key words:** academic success, competency-based learning, online education, assessment validity.

* Pascal Roubides, Fisk University, Nashville, TN, USA; proubides@fisk.edu; ORCID 0009-0005-8819-3533

## Introduction

Rapid expansion of online learning environments has intensified the need to critically examine how educators define and measure academic success (and this argument is true regardless of the modality of delivery). In higher education, success is often equated with quantifiable metrics such as performance grades, completion rates, retention, and "success" rates, and, without a doubt, test scores. However, these indicators may not accurately reflect genuine learning or intellectual growth. The challenge becomes particularly pronounced in disciplines such as mathematics, where conceptual understanding often develops nonlinearly and may not align neatly with performance outcomes.

This study emanated from the impetus to investigate student success factors in online learning environments, with multiple foci, such as the common metrics of retention and success, as defined by institutional standards, learner and instructor engagement, constructivist delivery approaches, infusion of Open Educational Resources (OER), contextualization, and others. One aspect of this research sought to determine whether a hybrid, polyparadigm competency-based model applied in online College Algebra courses could produce statistically significant improvements in retention and success rates compared to traditional courses based strictly on commercial resources, including related online assessment systems.

Neither commercial publisher resources, nor their assessment platforms were a focus of the investigation, it simply happened to be the "traditional" approach used for most similar courses at the researcher's institution at the time of the inquiry. As such, no conclusions can be drawn or expected about the effectiveness of any commercial resources or software systems.

## 1 Literature review

Academic success has traditionally been associated with measurable performance indicators, such as via various grading scales. Grades usually reflect measures of performance of what faculty or institutions perceive as valuable factors that denote academic achievement and are commonly used as measures of cognition (Borghans et al., 2016). Examples of activities that are commonly used as indicators of academic success include various graded assessments (quizzes, projects, examinations, etc.), homework, participation in class discussions, even attendance or class behavior. In other words, grading scales represent a composite score of student work, effort, achievement, perhaps personality traits, behavior, or simple adherence to mandated policies and procedures.

Brookhart et al. (2016) discuss arguments that despite any popularity of commonly used grading scales grades "are completely subjective measures of academic knowledge." The authors note that several studies in the early 20$^{th}$ century seemed to view class grades as simply unreliable. Other studies in recent years have attempted to illustrate relationships between grades and various cognitive and noncognitive factors representing student work or effort in class and/or individual traits that may impact learning or performance.

Despite any results from any particular study, educational researchers increasingly argue that these metrics fail to capture the full spectrum of learning experiences (Norris et al., 2008; Raffaghelli & Grion, 2023; Roubides, 2025). Quantitative assessments emphasize outcomes rather than processes (Mertens, 2019), often reducing complex cognitive and affective phenomena into simplified numerical representations, such as via arbitrary grading scales.

In order to address some of these limitations, one may turn to different educational models, such as competency-based education (CBE). This model focuses on demonstrated mastery of specific learning objectives rather than on a time-based progression (Walde, 2020). Though there are still many challenges with ensuring effective design and implementation of this educational approach (Gruppen et al., 2016), it is one way to address the issue with traditional grading caveats. Therefore, at least partial course redesign may be necessary in order to capture learning and learning growth more objectively.

In terms of course redesign, it is also possible to stray from the traditional choice of using commercially available course materials and instead turn to Open Educational Resources. Miao et al. (2019) describe OER as "teaching, learning, and research materials that utilize appropriate tools, such as open licensing, to permit their free reuse, continuous improvement and repurposing by others for educational purposes." Adil et al. (2024) posit that "information technology has changed the way of teaching and learning in academia" and that "open educational resources are quickly becoming significant components in higher education." The proliferation of present-day high quality OER materials further contributes to rethinking educational design by emphasizing accessibility, personalization, contextualization, learner autonomy, and of course they also alleviate in part the issue of affordability especially for marginalized cohorts and other learners typically struggling with the financial aspect of higher education (Lambert, 2020).

Integrating CBE with OER design, as was attempted in the present study, provides an opportunity to address a multiplicity of challenges all at once, even though addressing all such challenges is not the focus of the current investigation. This hybrid approach does however allow for the evaluation of not only student performance but also of the pedagogical frameworks that underpin learning, performance, and overall "success."

## 2 Methodology – Part 1

The study presented herein employed a quasi-experimental quantitative design comparing two groups of students of mathematics at a large, multi-campus, minority-serving state college in south Florida, with total enrolment of more than 51,000 students (Broward College, 2025). A very large percentage of these students (77.3%) are part-time students, have an average age of 24 years, with the majority of the students being female (57.9%), and designated as minority students (73.9% Hispanic/Latino or Black/African American) – in addition, more than half of enrolled students are first-generation college students, i.e., the first in their family to attend a post-secondary level institution (Broward College, 2025).

The two groups under study were: (1) an experimental group enrolled in a hybrid polyparadigm competency-based model of delivery, and (2) a much larger control group of historical data of students enrolled previously in similar traditional College Algebra courses. Both course types were offered in accelerated 8-week terms given the popularity of this short term for several courses including the one under investigation.

The sample size of the experimental group was 35 students, which is considered relatively high given that it was comprised of students in a single section of the target course during the 2024-2025 spring semester whose maximum allowable enrolment stood at 37 students. Data obtained from this section was compared to historical data from multiple sections using the traditional delivery approach at the college with a total sample size of 276 students.

Metrics of interest for this study included the rates of retention, which can be described as the proportion of students who did not withdraw from the course as a function of the course's total initial enrolment (i.e., grades issued must only include grades of A, B, C, D, or F), and the course success rates, which is a metric defined as the proportion of students enrolled in the course who were assigned a final course grade of A, B, or C.

The grade scale used in the course is the popular A-F letter grade scale, where A denotes "excellent performance" represented by the numerical range of [90-100%], B denotes "good performance" represented by the numerical range of [80-90%), C denotes "satisfactory performance" represented by the numerical range of [70-80%), D denotes "less that satisfactory performance" represented by the numerical range of [60-70%), and F denotes "unsatisfactory performance" represented by the numerical range of [0-60%).

A hypothesis test was set up to examine whether the difference of proportions of the two metrics (retention and success) between the two groups of data was negligible or significant as shown below:

$$H_o: \hat{p}_1 - \hat{p}_2 = 0$$
$$H_o: \hat{p}_1 - \hat{p}_2 \neq 0$$

where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions. The test statistic z is given by the formula

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left[\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]}}$$

where $\hat{p}$ is the pooled proportion calculated as

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

with $x_1$ and $x_2$ being the number of successes in each sample, and $n_1$ and $n_2$ being the sample sizes.

Conditions to conduct a valid two-proportion z-test were met, and those conditions included randomness and independence of sample data, data being dichotomous, and large enough sample sizes to ensure normality of the underlying distribution (Figure 1), and more specifically, ensuring that both $np \geq 10$ and $n(1-p) \geq 10$ for both samples (Agresti & Franklin, 2014). Subsequently, a two-sample, two-tailed z-test was conducted to evaluate the differences in the proportions of both metrics at a 95% confidence level.

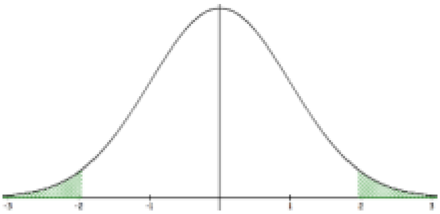

*Figure 1.* A graphical depiction of the expected normality of the distribution.

## 3 Results – Part 1

Data collected showed that in the experimental group comprised of 35 students, 21 students received grades of A, B, or C, representing a 60% "success" rate (21/35), 5 students received grades of D or F, representing a 14.3% "failure" rate (5/35), and 9 students withdrew from the course receiving a grade of W, representing a 25.7% "withdrawal" rate, or equivalently, a retention rate of 74.3%. Similar data collected from the control group comprised of 276 students, 130 students had received grades of A, B, or C, representing a 47.1% "success" rate (130/276), 65 students received grades of D or F, representing a 23.6% "failure" rate (65/276), and 71 students had withdrawn from the course receiving a grade of W, representing a 25.7% "withdrawal" rate, or equivalently, a retention rate of 74.3% which was identical to the retention rate in the experimental group.

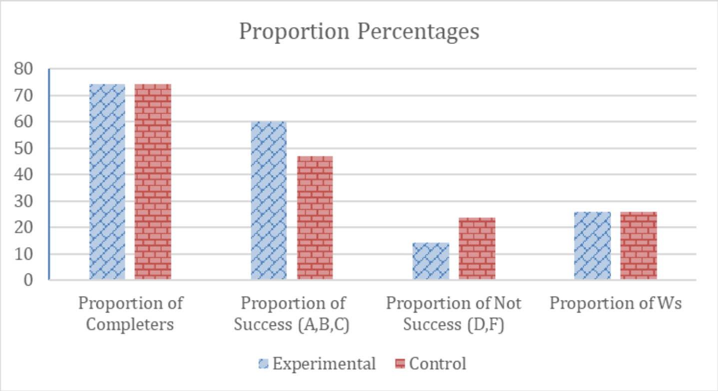


*Figure 2.* A graphical comparison of data from the two samples.

Figure 2 shows a graphical depiction of the descriptive statistics discussed for the two samples. Given the identical retention/withdrawal rate seen in both the experimental sample data and the control sample data, it was apparent that retention in the experimental group did not improve (nor decrease) in any significant manner, and as a formality only, confirmation of the hypothesis test for retention showed that the null hypothesis could not be rejected. Therefore, retention rates did not need to be analyzed or discussed further, and the focus of the analysis shifted solely to testing the proportions of “success” of the two samples.

Statistical analysis for the success proportions of the two samples showed that the p-value was 0.1503, larger than the 0.05 alpha level set for the test. Equivalently, the 95% confidence interval for the difference of the two proportions (-0.0468, 0.3048) included the assumed zero success proportion difference of the null hypothesis (Figure 3). This result indicated that the null hypothesis could not be rejected, in other words, there was no statistically significant difference in the success proportion of the two groups.

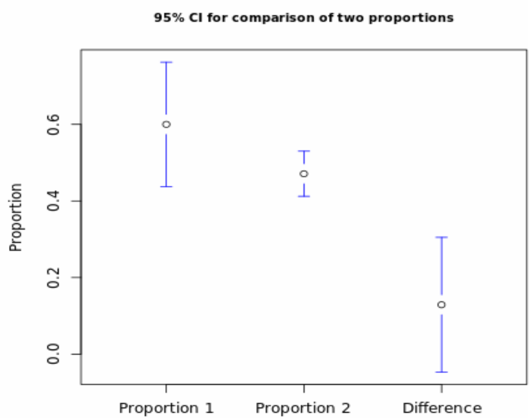


*Figure 3.* A 95% confidence interval plot for the success rates and their difference of the two groups being tested.

## 4 Discussion – Part 1

The results obtained from the hypothesis test conducted were rather surprising given the effort necessary to redesign the course in question so as to implement a much more complicated model of delivery than previously employed aimed at improving the success of students and were contrary to the researcher's expectations. A 60% success rate, even though representing raw data, appeared to signify a substantial improvement over the average success rate seen in historical data (47.1%) which was in line with empirical experiences of success rates in the course under investigation hovering around the 50% mark for many years. Hence, a question arose as to why this occurred.

It is not expected that statistical analyses provide perfect results or that they must confirm expectations, but it is possible that nuances with certain factors may have contributed to the results obtained. For example, the validity of the z-test may have been compromised due to the large sample size difference between the experimental group and the control group. Validity appears to be higher in samples of similar sizes than samples having large differences in sizes.

An inquiry into this potential using a mathematical approach for providing further insight as to whether the proportion difference between the two samples (0.129) may be considered small or large, can be seen in the related Cohen's *h* coefficient (Cohen, 1988). Cohen's *h* coefficient is defined as

$$h = 2\left|\arcsin\left(\sqrt{\hat{p}_1}\right) - \arcsin\left(\sqrt{\hat{p}_2}\right)\right|$$

and as a rule of thumb, *h* values at or below 0.2 indicate a small effect, *h* values from 0.2 to 0.5 indicate a medium effect, and *h* values from 0.5 to 0.8 indicate a large effect.

In the case under investigation, Cohen's *h* coefficient evaluates to 0.26, exceeding Cohen's suggested threshold of 0.2 for deciding that only a small effect exists, hence, it is possible that differences in sample sizes may play a role in the results obtained. Having said that, Cohen's suggested levels of the effect size are also rather arbitrary and based on empirical observations, and according to Cohen himself are "recommended for use only when no better basis for estimating the index is available" (Cohen, 1988). Therefore, this analysis can neither conclude nor refute possible issues with the validity of the test due to large differences in samples sizes of the two groups.

A different factor may have been at play in the results presented above, and based on prior empirical observations, a slightly different approach was taken to re-examine the data from this experiment as discussed in the sections that follow.

## 5 Methodology – Part 2

The methodology for the re-examination of data from this experiment was adjusted by altering the definition of "success" from the traditional definition of the proportion of students enrolled in the course who were assigned a final course grade of A, B, or C, to the proportion of "successful course completers" who were assigned a final course grade of A, B, or C.

The difference between the two definitions is based on the aim to move away from a definition of "success" as a function of initial enrollment, a measure which is more administrative than academic, and instead redefining "success" in a manner that is more conducive to describing academic performance intent, effort, and progress toward course completion. Successful course completers were defined as those students who completed at least two-thirds of the course and were assigned a final course grade of A, B, or C.

Students may formally withdraw from a course at the researcher's institution up to the $60^{th}$ percent of the enrolled term's duration and be assigned a grade of W, but as it is quite common, students may simply stop attending class with no formal withdrawal. In such situations, a student who simply stops attending class or stops completing any assigned class work the day after class has begun is still considered as "attending" class and is counted in the "success/failure" data under the traditional "success" definition. It is logical to argue that any student who stops attending the day after class has begun and never returns to class or never does any further work in class is impossible to complete the class successfully.

In the redefined definition of "success" as a proportion of successful course completers, a larger emphasis is placed on which students can actually be considered as "attending" class and putting on some effort towards completing the course. If a student "attends" class and completes some class activities, i.e., if there is an attempt or effort to learn enough and perform well enough to complete the course, then there is an opportunity that the student may perform well enough to be considered a "success." It is not possible for a student who is not attending class and does not complete any required work or does not put in any effort to complete the course to be considered a "success." In this sense, the redefined definition of "success" attempts to capture students' intent, effort, and progress toward completing the course, in addition to their academic performance in the course, as opposed to be simply counted as one of the many pieces of numerical data to be used to conduct mathematical tests on.

Hence, the new definition of "success" as those students who completed at least two-thirds of the course and were assigned a final course grade of A, B, or C was chosen to closely follow the timeline after which students who did not withdraw officially from the course were still counted in the data even though they may have already stopped attending class or doing any further class work. The "2/3 of the course" demarcation is still arbitrary and can be further adjusted if necessary but at least it does capture more students who have shown intent or have made an attempt or significant progress towards completing the course.

Conditions to conduct a valid two-proportion z-test under the new definition were mostly met, except for one measure; the conditions of randomness and independence of sample data, as well as the data being dichotomous were met. The conditions of large enough sample sizes to ensure normality of the underlying distribution, and more specifically, ensuring that both $np \geq 10$ and $n(1-p) \geq 10$ for both samples (Agresti & Franklin, 2014) were partially met. For the experimental group, the condition $np \geq 10$ was met, but n(1-p) evaluated to 7, i.e., less than the expectation that $n(1-p) \geq 10$, which may pose a concern. Given that the success proportion for the experimental group was 0.8, a skewness in the underlying distribution would be expected.

A related variation of the above normality conditions based on the "3-sigma rule", a common statistical advice (Wallis, 2013) that $np > 9(1\text{-}p)$ is also met. Other authors (Eusea et al., 2024) consider the condition that $np \geq 5$ and $n(1-p) \geq 5$ is sufficient to ensure normality of the underlying distribution, even though the larger the products, the better the approximation is expected to be. Given the non-exact nature of these conditions, it was decided to move forward with the two-sample, two-tailed z-test analysis at the same 95% confidence level as previously. In addition, in order to ensure that effect sizes did not alter results obtained, it was decided to conduct the same z-test accounting for a continuity correction, as well as conduct Fisher's exact test at the same confidence level as a second layer of validity of results obtained from the z-test analysis.

Everything else described in the methodology section above (Part 1) is still applicable and the same hypothesis test was set up to re-examine whether the difference of proportions of the redefined metric of success between the two groups of data was still negligible or significant.

## 6 Results – Part 2

The same data as described in the results section above (Part 1) were re-compared under the new definition of success. Based on the redefined metric, data collected showed that in the experimental group comprised of 35 students, 28 students who were still attending class at the 2/3 point of the 8-week term, received grades of A, B, or C, representing an 80% "success" rate (28/35). Similar data collected from the control group comprised of 276 students, 162 students who were still attending class at the two-thirds point of the 8-week term, had received grades of A, B, or C, representing a 58.7% "success" rate (162/276).

Statistical analysis for the success proportions of the two samples under the new definition of "success" showed that the p-value was 0.0149, which is this case is lower than the 0.05 alpha level set for the test. Equivalently, the 95% confidence interval for the difference of the two proportions (0.0684, 0.3577) did not include the assumed zero success proportion difference of the null hypothesis. This result indicated that the null hypothesis ought to be rejected, in other words, there was a statistically significant difference in the success proportion of the two groups.

As noted previously, secondary analyses were conducted to confirm or refute these results. A z-test with a continuity correction showed a p-value of 0.0244, not significantly different from the p-value obtained for the z-test without the continuity correction. Moreover, Fisher's exact test showed a p-value of 0.0163 which is also below the 0.05 threshold. These additional analyses provide confirmation that a statistically significant difference exists in the success proportions of the two groups analyzed.

## 7 Discussion – Part 2

The results obtained from the hypothesis test conducted using the redefined success metric were different than the results obtained from the same groups of data earlier. In other words, differences in successful completion rates between the groups were statistically significant, as concluded by both the z-test analysis and Fisher's exact test, suggesting that the experimental model positively influenced student success.

There is still some concern about the validity of the statistical analyses conducted which may have been compromised due to large sample size differences between the experimental group and the control group as well as the large success rate observed in the experimental group. As stated previously, validity appears to be higher in samples that are closer in size than samples having large differences in sizes. Cohen's *h* coefficient was calculated for the new proportions obtained under the redefined success metric and this time it evaluates to 0.47, still exceeding Cohen's suggested threshold of 0.2 for deciding that only a small effect exists and approaching the threshold for the assumed range indicating a large effect size. As determined previously, it is still possible that differences in sample sizes may play a role in the results obtained adding to the doubt whether there may be an issue in the validity of the tests conducted.

These statistical possibilities and confounding factors aside, the main conclusion that one may draw from the above analyses is that at least there is ambiguity in what we measure, how we measure it, and what that represents. The traditional metric of student "success" as a proportion of initial enrollment may be an inaccurate descriptor of true academic success. As such, different metrics accounting for a larger part of academic performance, intent, effort, and progress towards completing the course, may be necessary to capture the notion that students are or are not successful. Moreover, the underlying assumption that learning can be directly measured by performance on class activities may also be invalid. Would students be really successful if they earn a certain letter grade that is not an accurate descriptor of their level of learning?

## 8 Limitations and future research

As with every study, there are potential limitations in the results obtained in the study presented herein, as well as confounding factors some of which have been discussed in previous sections. In brief, this study's limitations include a rather small experimental sample, substantial differences in sample sizes between the two groups being compared, thus increasing the potential for decreased validity of the testing procedure, potential researcher bias, and confounding variables related to the course delivery design and implementation, even the actual metrics being measured.

Subsequent inquiry on this subject should consider employing mixed-method research approaches combining traditional quantitative analysis with qualitative data and metrics, possibly including student lived experiences, perceptions, reflections, etc. and related qualitative metrics to complement quantitative metrics. It can be argued that employing different research designs and broader metrics of several variables instead of complete reliance on arbitrary performance scales may result in a more holistic understanding of what academic success really entails and a better picture of how one can perceive success in academic settings.

## 9 Further discussion

Despite any limitations regarding the actual quantitative results of this study, the ambiguity of these results and their apparent dependence on the definitions of metrics used provides an important context in our understanding of academic success – current approaches do not appear to be able to capture the complexities of how success can be measured. New questions arise and these questions can only be addressed by a proper framework or even redefining the metrics being used for measuring such a subjective variable. In a business world analogy, would a business be considered successful if it were profitable but at the same time its employees were miserable, and the business operations destroyed the communities or environment in which it operated? Similarly, would students be considered successful if they received grades of A, B, or C but those grades did not directly correspond to the amount of learning having taken place?

It is worth noting that other than grades of A, B, or C, many or most U.S. institutions consider a grade of D as a "passing" grade earning college-level credit (Schneider & Hutt, 2023), but at the same time a grade of D is not accepted as part of the "success" metric. This is because D grades are considered as "unsatisfactory performance" in a course, and such low academic performance level is not considered sufficient for students to progress to any higher-level course requiring at least "satisfactory" performance to ensure sufficient academic preparation to move forward. D grades are also limiting student's eligibility to receive financial aid, including grants, loans, tuition reimbursement, or other similar financial aid, and therefore such grades are indirectly considered not to be part of "success" – however, they are not considered failing grades either (Scott-Clayton, 2013).

The commonly used U.S. letter grade performance scale is simply an arbitrary scale invented for administrative convenience, rather than based on any scientific basis (Penn State, 2023). It is also based on widely varied criteria and requirements from faculty to faculty, from course to course, and from institution to institution. As such, it is at least ambiguous that one can measure true student academic success based on such an arbitrary scale and be able to compare the "success" of one student versus the "success" of another.

That nuance aside, even with the commonly used arbitrary performance scale, had grades of D earned by students in the above experiment been included in the "success" rates (given that grades of D are still considered "passing" and that may be thought of as partial "success"), the overall success rates measured would probably have been different – even though it cannot be assumed that any statistical analyses would show that the experimental delivery approach employed as part of the above study would have had a positive or any more positive effect on student success.

There are many distinct factors that are contributing to our current inability (or perhaps unwillingness) to attempt to capture much more accurate levels of academic success. It is widely accepted that success is relative and as such it is subjective – the perception of success for any one individual (or institution, or organization) may be completely different than any other (Duckworth et al., 2012). Furthermore, academic success implies (or ought to imply) that a certain level of learning has been accomplished along with any performance metric achievement. To this end, it is important to note that learning itself is qualitative and chaotic (Roubides, 2015), and that the task to measure learning accurately is a particularly challenging task (Lopez, 2002; Cornell University, 2025). Educators usually rely on quantitative tools to measure this qualitative variable – this in itself is a limitation, and it is problematic. Also, there is a tendency to assume that measuring academic performance based on arbitrary criteria, arbitrary requirements, some of which have nothing to do with actual learning, captures learning growth and academic success objectively. Lastly, we assume that student "success" is related to initial class enrollment. This may prove to be one of the most important limitations for any studies on student "success" rates, such as the one presented herein.

## Conclusions

The discussion in this monograph addressed questions arising from a case study that had the partial aim to investigate the most prevalent metric of academic success in higher education. The impetus emanated from a quasi-experimental quantitative design comparing course outcomes from an accelerated online College Algebra course employing a polyparadigm delivery approach and historical outcomes from similar traditional courses whose design was based on commercial publisher resources and assessment platforms.

The findings of this study point to an ambiguity on the success of the new approach, which showed no statistically significant differences in the traditional definition of "success" as a function of total initial enrollment, but at the same time, positive results emerged when the "success" metric was interpreted as a function of successful course completers instead. Results discussed suggest that conventional performance-based metrics may not adequately represent authentic learning or student academic growth, including student intention and effort towards completing the course enrolled in. A broader reconceptualization of student academic success appears necessary, as well as how such a qualitative and chaotic variable can be measured properly in order to capture more precisely the level and essence of true academic performance.

## References

Adil, H. M., Ali, S., Sultan, M., Ashiq, M., & Rafiq, M. (2024). Open education resources' benefits and challenges in the academic world: A systematic review. *Global Knowledge, Memory and Communication*, *73*(3), 274-291.

Agresti, A., & Franklin, C. (2014). *Statistics. The Art & Science of Learning from Data*. Essex: Pearson Education.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354-13359.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of educational research*, *86*(4), 803-848.

Broward College. (2025). *Public Information Dashboard*. *Student Enrollment and Demographics.* Retrieved from https://www.broward.edu/about/

Cohen, J. (1988). *Statistical Power Analysis for the Social Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cornell University. (2025). *Measuring Student Learning*. Retrieved from https://teaching.cornell.edu/teaching-resources/assessment-evaluation/measuring-student-learning

Duckworth, A. L., Weir, D., Tsukayama, E., & Kwok, D. (2012). Who does well in life? Conscientious adults excel in both objective and subjective success. *Frontiers in Psychology*, *3*, 356.

Eusea, J., Okwan, P., Belmasrour, R., Patterson, S., & Andrus, S. (2024). *Introductory Statistics*. LOUIS Pressbooks.

Gruppen, L. D., Burkhardt, J. C., Fitzgerald, J. T., Funnell, M., Haftel, H. M., Lypson, M. L., ... & Vasquez, J. A. (2016). Competency-based education: Programme design and challenges to implementation. *Medical Education*, *50*(5), 532-539.

Lambert, S. (2020). Online open education and social justice: Progress for regional, multi-lingual, and female learners. *Inclusive Education Is a Right, Right*, 21-36.

Lopez, C. L. (2002). Assessment of student learning: Challenges and strategies. *The Journal of Academic Librarianship*, *28*(6), 356-367.

Miao, F., Mishra, S., Orr, D., & Janssen, B. (2019). *Guidelines on the Development of Open Educational Resources Policies*. UNESCO Publishing.

Mertens, D. M. (2019). *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*. Sage Publications.

Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review*, *43*(1), 42.

Penn State. (2023). *Do we need letter grades?* Retrieved from https://sites.psu.edu/mcmcivic.

Raffaghelli, J. E., & Grion, V. (2023). Beyond just metrics: For a renewed approach to assessment in higher education. In *Data cultures in higher education: Emergent practices and the challenge ahead* (pp. 89-121). Cham: Springer International Publishing.

Roubides, P. (2015). Distance learning and the control theory metaphor. *Procedia Computer Science*, *65*, 269-277.

Roubides, P. (2025, October). *Measuring Academic Success* (Beyond Reasonable Doubt) [Keynote address]. 8th World Conference on Education & Teaching, Lyon, France.

Schneider, J., & Hutt, E. L. (2023). *Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (but don't have to)*. Harvard University Press.

Scott-Clayton, J. (2013). Information constraints and financial aid policy. In *Student financing of higher education* (pp. 75-97). Routledge.

Walden, P. R. (2020, August). Competency-based education: Purposes and promises. *Seminars in Speech and Language, 41*(4), 289-297. Thieme Medical Publishers.

Wallis, S. (2013). Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, *20*(3), 178-208.