International Journal of Education, Modern Management, Applied Science & Social Science (IJEMMASSS)
ISSN : 2581-9925, Impact Factor: 6.340, Volume 04, No. 01(I), January - March, 2022, pp.01-08

COMPARING STUDENT'S PERFORMANCE IN HIGH SCHOOL ASSESSMENT SYSTEMS AND NATIONAL ASSESSMENT SYSTEMS FOR ADMISSION TO COLLEGES AND UNIVERSITIES IN SAUDI ARABIA

Khalid Al-Nowibet[*]
Shafiq Ahmed[**]
Latifah Al-Qasem[***]
Madhusudhan Prasad Varanasi[****]

ABSTRACT

This paper presents evaluation of assessment systems for sustainability of higher education in KSA. Previously, student's admission data to compare students' performance in the High School Assessment System (HSAS) and in the National Assessment System (NAS) were used (Alnowibet and Ahmad 2015). This study is further extended to investigate the level of performance in both male and female students using both assessment system's data in Saudi Arabia. A data set of university students specialized in selected specialties offered in Saudi universities is used. The outcome point out that there is a substantial student's performance difference in both assessment systems. Also, the results clearly shows that there is substantial performance difference between both categories; female students perform significantly better than their male counterparts. This research study can assist the Education policy makers to ameliorate the effectiveness of the High School Assessment System in Saudi Arabia for the sustainable development of the education system. The findings of the research may be useful for education policy formulation in Saudi Arabia.

Keywords: *High School Assessment System(HSAS), High School GPA (HSGPA), National Assessment System(NAS), Student Performance, Learning Outcomes, Aptitude Tests.*

Introduction

Education is an essential "driving force" of progress and it is necessary for the realization of many social, economic, political, and cultural benefits (Gao et al. 2016). Therefore, it is important for a country to design its education system wisely. In Saudi Arabia, for college admissions high school grades are normally viewed as unreliable while standardized tests such as national assessment test system are seen as methodologically rigorous for assessing student ability and achievement.

The practice of National Standardized Test in Saudi Arabia is considered relatively new. Students successfully completing different levels of education are accordingly certified and those completing secondary school education may take the public examination conducted by the Ministry of Public Education for admission to the colleges and universities(Al-Sadan 2000).

[G] Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia.
[**] Department of Industrial Engineering, College of Engineering, King Saud University, Riyadh 11451, Saudi Arabia.
[***] Department of Management, College of Business Administration, Al-Yamamah University, Riyadh 11451, Saudi Arabia.
[****] Professor, College of Business Administration, P.O.Box 45180, Riyadh 11512, Kingdom of Saudi Arabia.

To assess the student's abilities language and logical thinking is used by the NAS, such as:

- Reading comprehension.
- Logical relations.
- Problem-solving performance.
- Inferential skills.
- Intuitional capabilities.

The exam contains six segments. All questions are of multiple choices. The candidate selects the premier choice and mark their answer on the response sheet. The response sheets are of OMR (Optical Mark Reading) type, where a candidate has to mark his/her choices, or supplementary personal details, by filling circles marked on a pre-printed sheet. Then, the answer sheet is graded by a scanning machine with the help of computer programs. The main goal of the exam is to assess the candidates' academic talent through examinations and/or inventories. The examination is used as an admission tool to quantitatively answer the following questions:

- To what level does a student be endowed with certain skills?
- Does the applicant's aptitude make him/her eligible for admission to a certain study course?
- To which study course should an applicant be admitted?

Prior to 2001, admission to the Saudi Universities was based on marks obtained in the high school exam. The major criteria at present used in the selection procedure for admission are academic records, aptitude test and accomplishment test. In 2001 the Ministry of Higher Education introduced reforms and introduced National Assessment System(NAS). While NAS-I is conducted semiannually and tests the in depth knowledge of the given reading comprehension and some math problem-solving skills in the form of multiple choice questions, NAS-II examines cumulative scientific understanding of three-year high school science subjects - chemistry, biology, physics, mathematics and English. The weights for admission to colleges and Universities are 20-30% for high schools grades, NAS-I 30% and NAS-II 30% to 40%. The NAS-I and NAS-II tests are conducted universally under the direction of the National Center for Assessment in Higher Education(NCAHE 2011).

Literature Review

A number of research studies have indicated that students' assessment play a vital role to develop students learning capabilities. Student's assessment system serves the dual purpose; it measures not only student's continuous learning progress as they go but also a useful and an effective way to acquire useful data to inform instructors for their own instructional practices effectiveness and use assessments to understand what success looks like and how to do better in future(Stiggins 2002). Ramsden (Ramsden 2003)starts his discourse of the impact of assessment on learning with 'hidden curriculum' to support the primacy of assessment in students' perceptions. Bloxham and Boyd (Bloxham and Boyd 2007)advocate their argument that the assessment methodology of a specific course has a major effect on student action. Standardized exams endeavor to offer quantification of individual differences in as explicit ways as possible (Anderson 1975).

Historically, the application of a standardized test for assessing student learning capacities and comprehension become a common practice in many countries. In the US, the Scholastic Aptitude Test (SAT) recognized since 1930s contains individual tests for mathematics, critical reading, and writing and used as a metric for admissions(Lemann 2004). In 1959 American College Test (ACT) was introduced as a common admission test to inquire students interests and institutes that suit them(Evans 2013).

Another research study conducted in USA in 2009 on national standardized tests such as SAT and ACT, provides two reasons behind that: (1) the exams gives a typical metric to assess students with various credentials and high school preparation; (2) the tests predicts how well students will do in college (Evans 2013).

Because of high inconsistency in the quality of secondary education, it is a cumbersome task for an admission officer to judge the level of meticulousness of a high school curriculum by analyzing a student's transcript. The standardized exam score accordingly works as an approach to analyze students on a uniform scale. It serves to recognize students that might be underachievers in secondary school however have high potential for prevailing in college and students that may have gotten extraordinary results in secondary school yet not be very much prepared for the extra rigors of the college educational modules(Evans 2013).

Since many college-bound students get GPAs around 4.0, secondary school grades lose significance for distinguishing students for college admissions (Camara and Echternacht 2000). Postsecondary institutions that utilize the American College Testing (ACT) along with high school grades and other supporting proofs can take essential choices about applicants and entering students with more degree of trustworthiness and assurance (Readiness 2005). In addition, Knowledge acquired during employment have much weightage than their education content (Serrano et al. 2015) and aligning learning outcome according to management for their benefits weakens the outcomes potential to direct teaching and learning (Havnes and Prøitz 2016).

Based on the study to validate the assessment system of American early childhood education, many recommendations and suggestions were made to improve the assessment system (Goldstein and Flake 2016). A study investigated the change in test-taking inspiration inside a 2-h cognitive low-stakes test and its association with test results. It was revealed that in general, preliminary test-taking motivation was a superior indicator of test results than change in inspiration(Penk and Richter 2017). Bieler and McKenzie (Bieler and McKenzie 2017), presented their study on the concept of sustainability in the strategic plans of Canadian higher education institutions.

While there are several studies in the western context, the researcher could not come across significant studies in the context of Saudi Arabia. Therefore, based on literature, it can be concluded that an evaluation of Saudi Education Assessment system is an important step for the development of a sustainable education system.

Research Questions

RQ1. While measuring student's aptitude, is there any difference in HSGPA, NAS-I and NAS-II?

RQ2. Whether there is any association between gender and aptitude of students as measured in HSGPA, NAS-I and NAS-II.

RQ3. Whether there is any relation between gender and achieving learning objectives through NAS I and NAS II.

To address the research questions, in the present research, different assessment systems are compared using statistical analysis, and recommendations are made to improve the system.

Methodology

Based on the above research questions, the following objectives and hypotheses are formulated:

- **Objectives of the Study**
  - To examine the difference in the students aptitude and assessment system (HSAS, NAS-I, NAS-II) used.
  - To understand gender factors affecting the aptitude as measured through HSAS, NAS-I and NAS-II.
  - To study if there is any relation between gender of students and their performance in achieving learning objectives through NAS I and NAS II.
- **Hypotheses**

H_{01}: There is no significant difference in the student's level of aptitude and type of assessment system.

$H_{01.1}$: There is no significant difference in the students level of aptitude in the high school assessment system and NAS-I.

$H_{01.2}$: There is no significant difference in the students' level of aptitude in the high school assessment system and NAS-II.

$H_{01.3}$: There is no significant difference in the students' level of aptitude in NAS I and NAS II.

H_{02}: There is no significant difference in the students' level of aptitude in type of test and gender.

$H_{02.1}$: There is no significant difference in the students' level of aptitude in HSGPA and gender.

$H_{02.2}$: There is no significant difference in the student's level of aptitude in NAS I and gender.

$H_{02.3}$: There is no significant difference in the student's level of aptitude in NAS II and gender.

H_{03}: There is no significant difference in the students level of performance in achieving learning objectives through NAS I and NAS II and gender.

- **Sample**

A sample of 32522 candidates (18665 males, 13857 females) successful from high school system in KSA in December 2015is considered for the purpose of the study.

- **Data Collection**

Secondary data are collected from a Saudi Arabian University which is geographically distributed across the nation and has huge number of students from disciplines such as: Management, Social Sciences, Sciences, Engineering, Technology, Arts, Medical Sciences, Education, Law and Agricultural Science. The data analysis is carried out using SPSS software.

Analysis, and Results

Ideally, students in Saudi Arabia spend three years in high school. Student learning and capabilities during high school period are assessed by the cumulative grade point average during this period. Therefore, it is expected that the high school system provides a better assessment of student learning outcomes. In order to proceed with data analysis, descriptive summary statistics of the collected data is presented in Table 1.

Table 1: Descriptive Statistics of each Assessment System

Assessment	Mean	Standard Deviation	1st Quartile	Median	3rd Quartile	Range
HSGPA	93.86	5.88	91.41	95.56	98.15	41.33
NAS-I	75.24	8.31	70.00	76.00	81.00	100.00
NAS-II	70.63	10.37	64.00	71.00	78.00	84.10

Table indicates that mean HSGPS with standard deviation of 5.8 is higher than NAS-I and NAS-II.

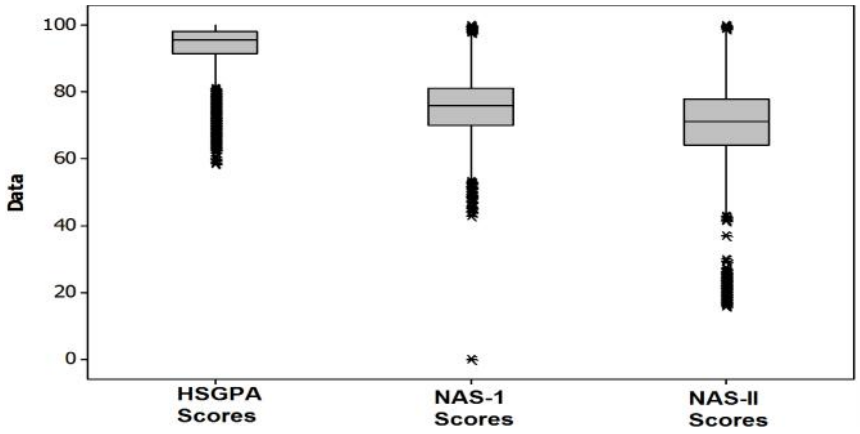

Fig. 1: Boxplot of High School GPA (HSGPA), NAS-I and NAS-II Scores

Boxplots reveal location and variation changes among data sets (Chambers et al. 1983). The three assessments systems were compared using the boxplot as shown in Figure 1. The inter-quartile ranges for HSGPA, NAS-I and NAS-II are 6.74%, 11% and 20% respectively, indicating NAS-I and NAS-II are better and more dependable compared to HSGPA for assessing student's learning attitudes and knowledge.

Table 2 presents the results of the comparison between HSGPA, NAS-I and NAS-II using two-sample t-value and *p*-value at 95% confidence level.

Table 2: Results of the comparisons between HSGPA, NAS-I and NAS-II using two-sample t-test

Samples	N	Ave. Diff.	Diff. 95% C.I.	*t*-value	*p*-value
HSGPA, NAS-I	32522	18.614	(18.503, 18.724)	329.99	~ 0.0001
HSGPA, NAS-II	32522	23.229	(23.099, 23.358)	351.58	~ 0.0001
NAS-I, NAS-II	32522	4.6148	(4.471, 4.759)	62.66	~ 0.0001

Since the p value in all cases is 0.0001< 0.05, all the null hypotheses $H_0$1 and respective sub hypotheses $H_0$1.1, $H_0$1.2, $H_0$1.3 are rejected.

In order to test hypotheses $H_0$2and $H_0$3, the set of data are separated according to gender to perform deeper analysis to investigate whether the above results hold for both genders. Table3 lists the descriptive statistics of both male and female in each assessment system. The descriptive statistics show that the difference in average score between both male and female is less than 1% in the HSGPA and NAS-I. Similarly, the difference between both genders in standard deviation is less than 1% in HSGPA and NAS-I. Table 3 shows a difference of 6.28% in average score between male and female in NAS-II. However, the difference in variability is very low (less than 1%). We conclude that the variability measures of scores are the same for both genders in all three assessments. We also conclude that male and female have the same level of general aptitude and performance in high school. However, female students are better than male students.

Table 3: Descriptive Statistics of each Assessment System with Respect to Gender

Assessment	Gender	N	Mean	Standard Deviation	1st Quartile	Median	3rd Quartile	Range
HSGPA	Female	13857	94.449	6.423	92.58	96.79	98.85	41.33
	Male	18665	93.414	5.391	90.85	94.69	97.42	40.33
NAS-I	Female	13857	75.303	8.759	70.00	76.00	81.00	100.00
	Male	18665	75.195	7.950	70.00	75.00	81.00	48.00
NAS-II	Female	13857	74.230	9.425	68.00	75.00	81.00	840
	Male	18665	67.951	10.221	62.00	68.00	74.00	84.10

Similarly, in Figure 2 boxplot is used to visually compare between the three assessments systems with respect to gender. Our finding from the descriptive statistics is also apparent in the boxplots. The boxplot of both genders in HSGPA share the same location and spread. This is also true in NAS-I scores. The difference between genders is clear from the graph in the NAS-II where the boxplot of male student is shifted below the female boxplot. In general, this indicates that, on an average, female students have better scientific knowledge than male.

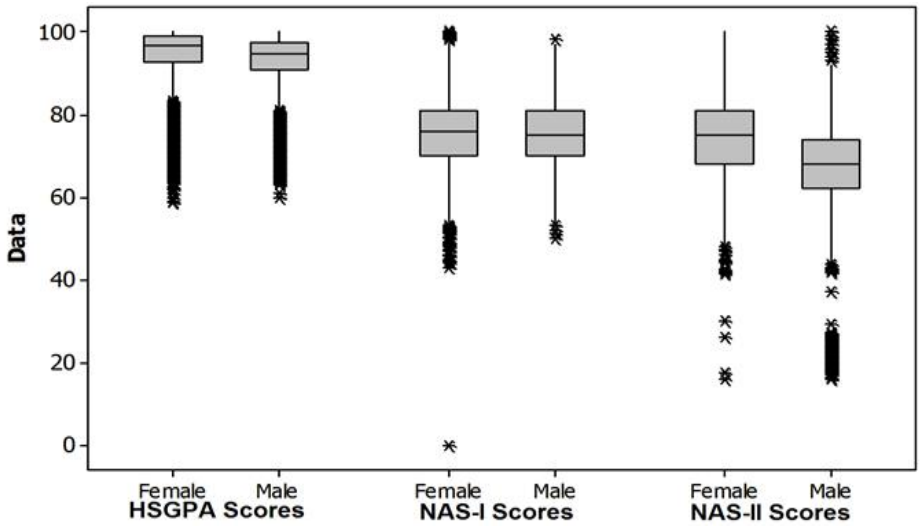

Fig. 2: Boxplot of high school GPA, NAS-I and NAS-II scores with Respect to Gender

Unpaired two-sample t-tests with respect to gender in all assessment systems were applied to measure the significance of the difference at 95% level. Table 4 lists the results of three different two-sample t-tests performed on each assessment. The first set of results in Table 4 represents testing the significance of the difference between male and female average HSGPA. Namely, the null hypothesis (H02) in the first set in Table 4 is that the difference between male's average HSGPA and female's average HSGPA is zero (H0 : μfemale – μmale = 0).

This means that there is a strong evidence of a difference between male's average HSGPA and female's average HSGPA. However, the magnitude of the difference is relatively small, within the interval (0.903, 1.167). Similar finding applies to the average scores of male and female students in NAS-II.

The results of *t*-tests in Table 4shows different findings for NAS-I. From Table 4, $H_0$2.2 is accepted with *p*-value (0.225), meaning, there is an evidence that there is no difference between male's average NAS-I score and female's average NAS-I score. This translated as the general aptitude of male students and female students are the same, yet female students' performance, in an average is better in acquiring scientific knowledge.

Table 4: Results of the Comparisons between Male and Female using Two-sample t-test for HSGPA, NAS-I and NAS-II

Samples	Gender	Ave. Diff.	Diff. 95% C.I.	*t*-value	*p*-value
HSGPA	Female	1.035	(0.903, 1.167)	15.38	~ 0.0001
	Male				
NAS-I	Female	0.107	(-0.078, 0.293)	1.14	0.255
	Male				
NAS-II	Female	6.279	(6.064, 6.494)	57.30	~ 0.0001
	Male				

$H_0$2.1 and $H_0$2.3 are rejected because their *p*-value is <0.0001 and as the *p*-value is 0.255 greater than 0.0001, $H_0$2.2 is accepted. This implies that there is significant difference in the student's level of aptitude in HSGPA, NAS-II and gender. However, there is no significant difference in the student's level of aptitude in NAS-I and gender.

To measure the performance of each gender in both NAS-I and NAS-II, a paired two-sample *t*-tests is performed in Table 4 with significance level of 95%. The first set of results in Table 5 represents testing the significance of the difference between female's average scores in NAS-I and NAS-II. The null hypothesis ($H_0$3) is then defined as the difference between female's average scores in NAS-I and NAS-II is zero ($H_0$3: female $\mu_{NAS\text{-}I}$ – female $\mu_{NAS\text{-}II}$ = 0). Similarly, the null hypothesis of the second set in Table 4 is $H_0$3: male $\mu_{NAS\text{-}I}$ – male $\mu_{NAS\text{-}II}$ = 0.

Table 5: Results of the Comparison for Male and Female Performance in NAS-I and NAS-II

Samples		Ave. Diff.	Diff. 95% C.I.	*t*-value	*p*-value
Female	NAS-I	1.073	(0.859, 1.287)	9.82	~ 0.0001
	NAS-II				
Male	NAS-I	7.244	(7.058, 7.430)	76.43	~ 0.0001
	NAS-II				

Results in Table 5show that *$H_0$3: There is no significant difference in the students level of performance in achieving learning objectives through NAS I and NAS II and gender* is **rejected** with very small *p*-value (<0.0001). It is obvious that the magnitude of the difference between female scores, within (0.859, 1.287), is much less than the magnitude of the difference between the male score, within (7.058, 7.430).

Conclusion

This research study has investigated the research questions and examined the reliability of the high school assessment system vs. the NAS-I and NAS-II. Results in the boxplots have clearly indicated that the location of high school assessment system's (HSGPA) data are totally dissimilar as compared to the NAS-I and NAS-II. If the HSGPA results are viewed as dependable in evaluating the fundamental aptitude and knowledge for students, in the inter-quartile range of the data the boxplot should have overlapped in location with the boxplots of NAS-I and NAS-II. In contrast, comparing between NAS-I and NAS-II boxplots, both NAS-I and NAS-II share almost half of the inter-quartile range. Thus, providing a strong evidence of superior and more dependable assessment systems in assessing student's learning skills and knowledge. Finally the research study has clearly indicated that there is significant performance difference between both genders and female students perform significantly better than male students.

Similar to the findings of Camara and Echternacht, 2000, as many high school students in Saudi Arabia are getting GPAs close to 4.0, admissions test results like NSA I/NAS II, and other info attained significance in college admissions. As mentioned in the review of literature, this finding is supported by Evans 2013 and Readiness 2005.

Before we can solve the challenges of assessment, we must agree on what we want students to know and then design tests that will measure their performance(Wiggins 1990). The effectiveness of the teacher is the major determinant of student academic progress. Hence as a part of viable educational assessment framework, Saudi Arabia need to concentrate more on connecting teacher adequacy to

student outcomes. The change in assessment practices can't be proficient in separate tests and measurement courses, but instead ought to be a chief concern in teaching methods courses (Shepard 2000). Formative assessment is a vital constituent of classroom work and can ameliorate student achievement (Black and Wiliam 2010). If we wish to upgrade student achievement, we should give careful consideration to the enhancement of class room assessment(Stiggins 2002).

Recommendations

This research study suggests that the high school system (HSGPA) should put more emphasis on the skills. One may argue: why not consider that NAS-I and NAS-II are under estimating the performance of the students. NAS-I and NAS-II are actually designed to measure some basic skills and knowledge that the student should have acquired during his study based on the learning objectives of the high school. However efforts needs to be made to ensure that the variation between HSGPA, NAS-I and NAS-II is minimum. This is possible through training programs that add value to teachers and by linking teacher effectiveness to the student outcomes, so as to make the education system more sustainable. Future research studies may focus on these and allied issues.

Acknowledgments

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding this Research Group NO (RGP-1436-040). Also, the authors are thankful to the Department of Information and Statistics at King Saud University for providing the data needed for this analysis.

References

1. Al-Sadan, I. A. (2000). Educational Assessment in Saudi Arabian Schools. *Assessment in Education: Principles, Policy & Practice, 7*(1), 143-155, doi:10.1080/713613320.
2. Alnowibet, K., & Ahmad, S. (2015). Saudi School Assessment System for Predicting Admissions to Science Colleges. *The Online Journal of Quality in Higher Education, 2*(1), 37-41.
3. Anderson, L. W. (1975). Student involvement in learning and school achievement. *California Journal of Educational Research, 26*(2), 53-62.
4. Bieler, A., & McKenzie, M. (2017). Strategic Planning for Sustainability in Canadian Higher Education. *Sustainability, 9*(2), 161.
5. Black, P., & Wiliam, D. (2010). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 92*(1), 81-90, doi:doi:10.1177/003172171009200119.
6. Bloxham, S., & Boyd, P. (2007). *Developing Effective Assessment In Higher Education: A Practical Guide*. UK: McGraw-Hill Education.
7. Camara, W. J., & Echternacht, G. (2000). The SAT® I and High School Grades: Utility in Predicting Success in College. (pp. 12). USA: College Board.
8. Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical Methods for Data Analysis (Statistics)*: Chapman and Hall/Cole Publishing Company.
9. Evans, B. (2013). *College Admission Testing in America* (International Perspectives in Higher Education Admission Policy: A reader). USA: New York: Peter Lang Publishing.
10. Gao, Y., He, Q., Liu, Y., Zhang, L., Wang, H., & Cai, E. (2016). Imbalance in Spatial Accessibility to Primary and Secondary Schools in China: Guidance for Education Sustainability. *Sustainability, 8*(12), 1236.
11. Geiser, S., & Maria Veronica, S. (2007). Validity Of High-School Grades In Predicting Student Success Beyond The Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. (pp. 35). USA: Center for Studies in Higher Education, UCF, Berkley.
12. Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability, 28*(3), 273-293, doi:10.1007/s11092-015-9231-8.
13. Havnes, A., & Prøitz, T. S. (2016). Why use learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educational Assessment, Evaluation and Accountability, 28*(3), 205-223, doi:10.1007/s11092-016-9243-z.

14. Lemann, N. (2004). A history of admissions testing. In R. Zwick (Ed.), *Rethinking the SAT : the future of standardized testing in university admissions*. USA: Routledge.

15. NCAHE (2011). General Aptitude Test.

16. Penk, C., & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation and Accountability, 29*(1), 55-79, doi:10.1007/s11092-016-9248-7.

17. Ramsden, P. (2003). *Learning to Teach in Higher Education* (2ed.): Routledge.

18. Readiness, C. (2005). Are High School Grades Inflated? Issues in College Readiness. *ISSUES IN COLLEGE READINESS*.

19. Serrano, G., Llamazares, F., & Otamendi, F. (2015). Measurement and Sustainability of the Qualifications Frameworks in the European Higher Education Area through an Employment Survey on Access to the Labour Market. *Sustainability, 7*(10), 13777.

20. Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher, 29*(7), 4-14, doi:doi:10.3102/0013189X029007004.

21. Stiggins, R. J. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan, 83*(10), 758-765, doi:doi:10.1177/003172170208301010.

22. Wiggins, G. (1990). The Case for Authentic Assessment. *Practical Assessment, Research and Evaluation, 2*(2), 1-3.

❖◆❖