F E A T U R E

A R T I C L E

Predicting NCLEX-RN Success

The Seventh Validity Study
HESI Exit Exam

ANNE YOUNG, EdD, RN
PAMELA WILLSON, PhD, RN, FNP-BC, CNE

The findings of six previously conducted studies indicated that the HESI (E^2) was highly accurate in predicting NCLEX-RN success. The purpose of this study—the seventh study to investigate the validity of the E^2—was to examine the accuracy of three parallel versions of the E^2 in predicting licensure success and to describe program practices regarding E^2 benchmark scores, remediation programs, and retesting policies. The findings of this study again indicated that the E^2 was highly accurate in predicting NCLEX-RN success. Additionally, all three versions of the E^2 were found to have a predictive accuracy above 90%. The most common E^2 benchmark score designated by faculty at the participating schools was 850, and most schools required students to retest with different versions of the E^2 until the faculty-designated E^2 benchmark score was achieved. Remediation seemed to be effective in raising students' E^2 scores, and it was recommended that future research investigate the effectiveness of specific remediation strategies.

K E Y W O R D S

HESI • NCLEX-RN • Validity

Six previously conducted studies (S1, S2, S3, S4, S5, and S6) examined the accuracy of Elsevier's HESI Exit Exam (E^2) in predicting NCLEX-RN success, and the findings of all six studies indicated that the E^2 was highly accurate in predicting licensure success.[1–6] Morrison et al[7] reported that research designed to quantify the degree of validity of HESI examinations is an ongoing process and further described criterion-related validity as inferences made from analyses of test scores for the purpose of predicting student outcomes on another criterion of interest, such as NCLEX-RN success. The findings of all six studies that investigated the accuracy of the E^2 in predicting licensure success provided evidence of the E^2's criterion-related validity.[7] Consequently, the authors of S5 and S6 concluded that the E^2 could be used confidently to assess students' preparedness for the licensure examination.[5,6]

The authors of S1 reported that when the administration of the E^2 was monitored, it was significantly more accurate in predicting NCLEX-RN success than when it was not monitored. The authors of S2 reported that low-scoring E^2 students were more likely to fail the NCLEX-RN than high-scoring students and that schools that used E^2 scores as a benchmark for remediation had significantly fewer low-scoring students who failed the NCLEX-RN than schools that did not use E^2 scores as a benchmark for remediation.[2] The S3 researchers asked deans and directors of the participating schools if the E^2 was used as a benchmark for remediation, and findings indicated that 52.15% of the students who were remediated using data provided by the E^2 passed the NCLEX-RN, whereas 47.85% of those who were not remediated using the E^2 as a guide for remediation failed the licensure examination. The authors of S3 recommended that future studies should obtain more information regarding how remediation was defined and implemented.[3] Consequently, the S4 authors collected data regarding the types of remediation strategies implemented by participating schools, thus providing a qualitative approach for defining remediation.[4] Nibert et al[8] investigated the use of the E^2 as a benchmark for progression and as a guide for remediation and concluded that using E^2 scores as benchmarks for progression was effective in providing a guide

for remediation, which enabled faculties to better assist students in completing the nursing curriculum and becoming successful first-time NCLEX-RN candidates. The authors of S4 and S5 also investigated the number of E^2 failures by scoring categories, and both studies reported that as students' E^2 scores decreased, NCLEX-RN failures increased.[4,5]

To evaluate the effectiveness of remediation, schools often administer a parallel version of the E^2 following the students' efforts to address weaknesses described in their individual E^2 scoring report. The sixth validity study examined the predictive accuracy of three parallel versions of the E^2, and the authors reported that there was no significant difference in the predictive accuracy between Version 1 (V-1) and Version 2 (V-2) of the E^2 but that Version 3 (V-3) was significantly less accurate in predicting NCLEX-RN success than V-1 and V-2. To validate the findings of S6, the authors of this study (S7) decided to further investigate the validity of parallel testing. Specifically, the purpose of S7 was to examine the accuracy of three parallel versions of the E^2 in predicting licensure success and to describe program practices regarding E^2 benchmark scores, remediation programs, and retesting policies.

Conceptual Framework

Classical test theory described by Crocker and Algina[9] and critical thinking theory described by Paul[10] were used as the theoretical basis for developing a model for writing critical thinking test items as described by Morrison et al.[11] Assumptions of classical test theory provide a mechanism for assessing error to ensure greater precision in the measurement of outcomes. Paul's definitions of critical thinking act as a guide to direct examination development that requires students' application of nursing judgment to select the correct examination response from highly plausible choices. This model was used to formulate the conceptual framework for this study, as well as for the six previously conducted studies that examined the accuracy of the E^2 in predicting NCLEX-RN success.

▇ INSTRUMENTATION

Two instruments were used to obtain data for this study: the E^2 and the Participant School Survey. The E^2 is a 160-item comprehensive examination, which includes 10 pilot items that do not contribute to the students' scores and are included in the examination for the purpose of obtaining item analysis data. The examination is usually administered during the final semester, or quarter, of the nursing curriculum. Scores range from 0 to approximately 1800, with the highest score dependent on the difficulty level of the test items included in the examination. Those who score 900 and above are described by Elsevier, the producer of the E^2, as predicted to pass the NCLEX-RN. The E^2 administered during the study mirrored the NCLEX-RN test plan, described by the National Council of State Boards of Nursing (NCSBN), and also emulated the test item formats the NCSBN implemented at that time. These test item formats included multiple-choice, multiple-response, fill-in-the-blank, hot spot, chart/exhibit, and drag-and-drop items.[12]

The reliability of the E^2 is determined by calculating the Kuder-Richardson formula 20 (KR20) for every examination that is returned to Elsevier for scoring. The estimated reliability coefficient, or KR20, for the E^2 is a measure used to estimate the reliability of these examinations prior to their administration. The calculation of the estimated KR20 is based on data obtained from all prior administrations of the test items included in the examination. The estimated KR20s for the three versions of the E^2 that were used to obtain data for this study were 0.91 for V-1, 0.98 for V-2, and 0.91 for V-3.

The validity of the E^2 has been established by the six previously conducted studies that investigated the accuracy of the E^2 in predicting NCLEX-RN success. Based on data obtained from a total sample size of 37 184 students from associate degree (ADN), baccalaureate (BSN), and diploma schools of nursing, the E^2 was found to be between 96.36% and 98.30% accurate in predicting NCLEX-RN success, and the predictive accuracy of the E^2 was not significantly different among the six previously conducted studies. Additionally, all six of these studies reported that the predictive accuracy of the E^2 was not significantly different among nursing programs: ADN, BSN, and diploma.[1–6]

The Participant School Survey was an electronic survey adapted from the S6 survey and consisted of 15 multiple-choice items with open comment fields. The purpose of this survey was to obtain information about (1) participating schools' policies regarding the use of the E^2 as a benchmark in determining students' preparedness for the NCLEX-RN, (2) remediation strategies implemented to assist students in achieving E^2 benchmark scores, and (3) students' outcomes on their first attempt at taking the NCLEX-RN.

▇ METHODOLOGY

The six previously conducted studies surveyed all programs that administered the E^2 during the study period and returned students' scores to Elsevier for analysis. However, the response rate for S6 was lower (32.19%) than the previous five studies, and the authors of S6 attributed this decreased response to participating deans'

Table 1

Description of Sample by Program Type

Program Type	Schools, n (%)	Students, n (%)
ADN	45 (62.50)	2557 (58.34)
BSN	22 (30.56)	1617 (36.89)
Diploma	5 (6.94)	209 (4.77)
Total	72 (100)	4383 (100)

and directors' reluctance to share data about students who failed the NCLEX-RN. The S6 authors therefore recommended that future investigators consider using a random sampling technique to survey the population. They explained that perhaps a smaller, representative sample could receive additional follow-up and assistance in completing the survey, as well as personal reassurances regarding security of the data and anonymity of the participating subjects.

The robustness of the S7 population made stratified random sampling an effective research technique. To determine a representative sample size for the population, a power analysis was conducted. An a priori sample size of 4336 students was determined to be representative of the population at a 95% confidence level. The total sample size for this study consisted of 4383 students and was therefore considered representative of the population. The sample was stratified by the three program types—ADN, BSN, and diploma—to match the proportion of students who took the NCLEX-RN during the 2007 calendar year. Following approval from the institutional review board, the electronic survey was sent to the deans and directors of 137 randomly selected schools of nursing that administered the E^2 between September 1, 2006, and August 31, 2007. Schools that did not respond within 2 weeks were sent an electronic reminder, and those that did not respond following the first reminder were sent a second reminder 2 weeks later. Reassurance regarding the removal of all identifying data was provided to the deans and directors who expressed concern about sharing student information. After identifying which scores were associated with NCLEX-RN failure, deans and directors were encouraged to remove student names before uploading the data.

■ FINDINGS

Of the 137 schools surveyed, 72 (52.55%) uploaded their responses for analysis. Participants were asked to identify first-time NCLEX-RN failures and to answer 15 questions regarding their school's benchmark scores, remediation programs, and retesting policies. To ensure confidentiality, questionnaires were returned electronically and all personal and school identifiers were removed prior to adding the information to the database. There-

fore, researchers who had access to the data could view aggregate data only. The sample consisted of 45 (62.50%) ADN programs, 22 (30.56%) BSN programs, and five (6.94%) diploma programs. The total student sample consisted of 4383 participants: 2557 (58.34%) ADN students, 1617 (36.89%) BSN students, and 209 (4.77%) diploma students (Table 1).

Accuracy of E^2 Parallel Forms

Predictive accuracy represents the ability of one testing outcome to predict success on another measure, or in this case, the accuracy of the E^2 in predicting success on the NCLEX-RN. Because those who score 900 and above on the E^2 are described by Elsevier as predicted to pass the NCLEX-RN, the predictive accuracy of the E^2 was determined by tabulating the number of students who scored 900 and above on the E^2 and calculating the percentage of those students who passed the licensure examination on their first attempt. Of the 1075 students who scored 900 and above on V-1, total implies the total group 1066 (99.16%) passed the NCLEX-RN on their first attempt. Students who did not achieve the faculty-designated benchmark score for their school were required to remediate and retest with a different version of the E^2. Of the 4383 total student sample, 730 (16.66%) students were required to take V-2 of the E^2, and 271 of these students scored 900 and above on the second version of the E^2. Of the 271 who scored 900 and above on V-2, of 259 (95.57%) passed the NCLEX-RN on their first attempt. Of the 730 students who were required to take V-2, 367 (50.27%) students were required to take V-3 of the E^2, and 148 of these students scored 900 and above on the third version of the E^2. Of the 148 students who scored 900 and above on V-3, 138 (93.24%) students passed the NCLEX-RN on their first attempt. A χ^2 was performed to determine if the predictive accuracy was significantly different among the three versions of the E^2 administered to the study subjects. Findings indicated that significantly more students who scored 900 and above on V-1 of the E^2 successfully completed the NCLEX-RN on their first attempt than those who scored 900 and above on V-2 and V-3 of the E^2 ($\chi^2(2) = 31.4156$, $P \leq .000$; Table 2). A

Table 2

Predictive Accuracy of E^2 by Versions[a]

Version	Scoring 900 and Above, n	Passed NCLEX-RN, n	Accuracy, %
V-1	1075	1066	99.16
V-2	271	259	95.57
V-3	148	138	93.24

[a]The predictive accuracy was significantly different among the three versions of the E^2 ($\chi^2(2) = 31.4156$, $P \leq .000$).

Table 3

No. of Participants and Predictive Accuracy Rate of E^2 Version 1 by Study

Study	Participants, n	Accuracy Rate, %
1	2555	97.28
2	3296	96.36
3	5588	97.64
4	5903	98.30
5	9695	97.83
6	10 147	96.44
7	4383[a]	99.16

[a]Study 7 used a random sample, whereas the previous six studies investigated the entire population.

total of 1494 students scored 900 and above on one of the three versions of the E^2, and 1463 (97.93%) students passed the NCLEX-RN on their first attempt. Therefore, the accuracy of the E^2 in predicting NCLEX-RN success was 97.93%, regardless of whether the student was required to take the examination up to three times before achieving the faculty-designated E^2 benchmark score.

Only S6 and S7 examined the predictive accuracy of parallel versions of the E^2. Therefore, when determining if the findings of this study differed from the six previously conducted studies, only V-1 findings were examined. The predictive accuracy of V-1 of the E^2 ranged from 96.36% for the second study to 99.16% for this study. A χ^2 was performed to determine if the predictive accuracy of the E^2 was significantly different among the seven studies. Findings indicated that there was a significant difference in the predictive accuracy of the E^2 among the seven studies ($\chi^2(6) = 45.179$, $P \leq .000$; Table 3).

Benchmark Scores, Remediation Programs, and Retesting Policies

Participants were asked if they had established policies that used the E^2 as a benchmark for remediation. Of the 72 participating schools that returned questionnaires for analysis, 67 (93.06%) schools responded to the question, and 48 (71.64%) of these 67 schools reported that they had established a benchmark score for the E^2. A score of 850 was established as a benchmark by 27 (56.25%) schools, 875 by 2 (4.17%) schools, 900 by 13 (27.08%) schools, and 950 by 3 (6.25%) schools. One school (2.08%) reported using 725, one school (2.08%) reported using 750, and one school (2.08%) reported using 800 (Figure 1). Because most schools used 850 as their benchmark, the predictive accuracy of the E^2 was calculated for those who scored 850 and above instead of 900 and above on the three versions of the E^2. Of the 2198 students who scored 850 and above on one of the three versions of the E^2, 2129 (96.86%) passed
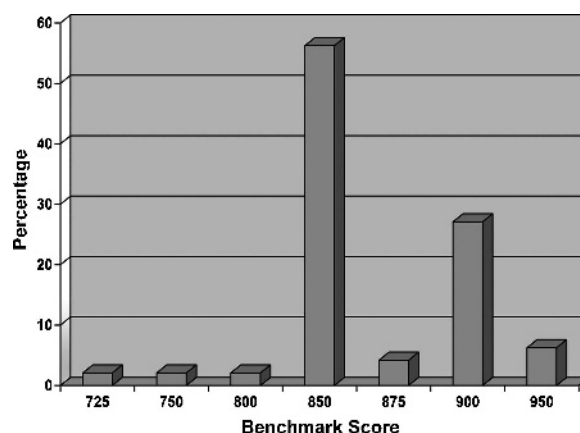

FIGURE 1. Percentage of schools designating specific E^2 benchmark scores.

the NCLEX-RN on their first attempt. Therefore, these findings indicated that the E^2 was 96.86% accurate in predicting NCLEX-RN success for students scoring 850 and above, regardless of whether they required retesting up to three times before achieving a score of 850.

Deans and directors were asked to describe the consequences established at their school for students who did not meet the faculty-designated E^2 benchmark score. Of the 66 schools that responded to this question, 45 (68.18%) schools reported having at least one consequence for failing to meet the faculty-designated E^2 benchmark score. Of the 45 schools that reported having at least one consequence for failing to meet the benchmark score, 24 (53.33%) required students to retake the E^2, a total of 21 (46.67%) delayed or denied NCLEX-RN candidacy, 20 (44.44%) delayed or denied graduation, 15 (33.33%) resulted in failure of a course, and eight (17.78%) resulted in failure of a capstone course (Figure 2).

Of the 45 schools that required remediation, Elsevier online examination remediation was used by 26 (57.78%) schools; the Elsevier NCLEX-RN Review Manual, by 25 (55.56%); other review books, by 18 (40.00%); remedial courses, by 16 (35.56%); computer-based tutoring,
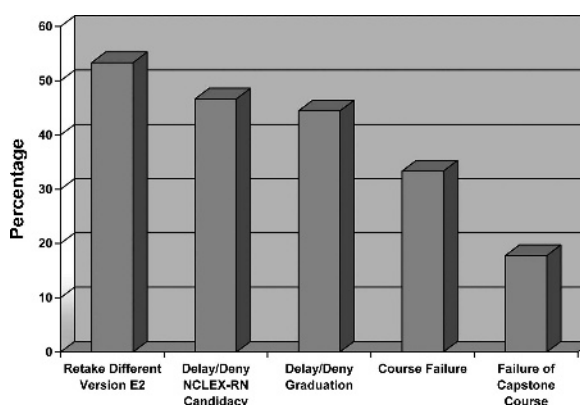

FIGURE 2. Consequences implemented by schools for students who did not meet the faculty-designated E^2 benchmark score.
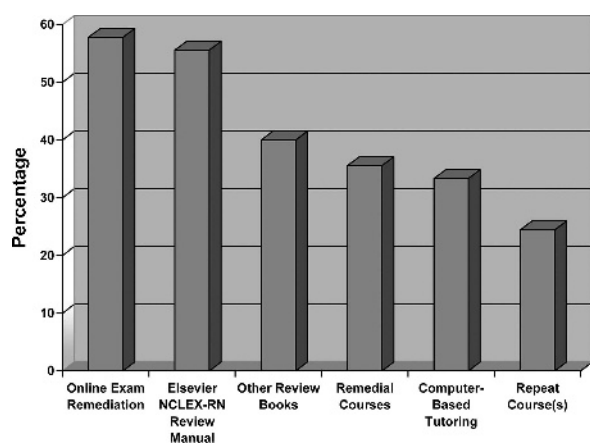
FIGURE 3. Remediation strategies implemented by participating schools.

by 15 (33.33%); and repeating courses, by 11 (24.44%) (Figure 3).

To evaluate the effect of students' remediation efforts, many schools required retesting with a different version of the E^2 following remediation, some until students achievedthe faculty-designated E^2 benchmark score. Following remediation, 44 (97.78%) respondents reported that retesting with a parallel version of the E^2 was required if students did not achieve the faculty-designated E^2 benchmark score. Of the 44 schools that required retesting, one retest was required by seven (15.91%) schools; two retests, by seven (15.91%); three retests, by 17 (38.64%); four retests, by four (9.09%); and more than four retests, by nine (20.45%).

■ DISCUSSION

This was the seventh study that investigated the accuracy of the E2 in predicting NCLEX-RN success, and like the previous six studies, the E^2 was found to be highly accurate in predicting licensure success. Based on recommendations of the S6 authors, a random sample was surveyed for this study rather than the entire population. The smaller sample size enabled the researchers to follow-up with nonresponders and provide additional assurance to deans and directors who were concerned about the confidentiality of student data. This strategy seems to have been effective because the response rate for S7 was higher by more than 20% from that of S6.

The first version of the E^2 for this study was found to be 99.16% accurate, the highest predictive accuracy of all seven studies. For the first time, a significant difference was found in the predictive accuracy of V-1 among all studies conducted. This finding is likely because of the fact that the predictive accuracy for V-1 of S7 was 99.16%, whereas the predictive accuracy of S2 was 96.36% and that of S6 was 96.44%. Although

these differences were significant, the findings of the seven studies nevertheless indicate that the E^2 is highly accurate in predicting NCLEX-RN success, as indicated by the range in predictive accuracy of 96.36% to 99.16% in the seven studies conducted with a total student sample size of 41 567.

The sixth and seventh validity studies were the first to investigate the predictive accuracy of two additional versions of the E^2 administered to those who did not meet their school's designated benchmark score. Like S6, a significant difference was found in the predictive accuracy by versions. However, unlike S6, in which V-3 of the E^2 was 82.50% accurate in predicting NCLEX-RN success, the predictive accuracy of all three versions of the E^2 for S7 was above 90%, indicating that for this study, the E^2 was highly predictive of NCLEX-RN success regardless of whether the student was required to take up to three versions of the E^2 before achieving the faculty-designated E^2 benchmark score.

Most schools (56.25%) designated 850 as their E^2 benchmark score, followed by 900 (27.08%). Interestingly, the predictive accuracy of the E^2 for those scoring 850 and above—regardless of whether students were required to take up to three versions of the E^2 before achieving a score of 850 and above—was 96.86%. Therefore, faculties that designate 850 and above for their school's benchmark score can be confident that their students are likely to pass the NCLEX-RN on their first attempt. Elsevier describes those who score 900 and above as predicted to pass the NCLEX-RN. Based on the findings of this study, 97.93% of those who scored 900 and above passed the NCLEX-RN on their first attempt, regardless of whether they were required to take the E^2 up to three times before achieving a score of 900 and above. Therefore, the score described by Elsevier, the producer of the E^2, as predicted to pass the NCLEX-RN, is a highly accurate predictor of NCLEX-RN success.

More than half of the schools that required remediation (57.78%) used the online examination remediation that is provided at the completion of the E^2. This remediation consists of test item rationales that explain why one option is correct and the other options are incorrect. The Elsevier NCLEX-RN review manual and other review books were described by participants as the next most frequently used remediation resources.

■ RECOMMENDATIONS

Measurement of validity is an ongoing process, and as such, investigations regarding the validity of the E^2 should continue to be conducted. Because of the robustness of the population of E^2 users, future studies should continue to use a random sample to obtain the data. This study and S6 investigated the predictive accuracy of

V-1, as well as the predictive accuracy of two additional parallel versions of the E^2 in predicting NCLEX-RN success. Both studies found that there was a significant difference in the versions' predictive accuracy. However, all three versions in S7 had a predictive accuracy above 90%. Because there was more than a 10% difference in the predictive accuracy of V-3 in S6 (82.50%) and V-3 in S7 (93.24%), this study should be replicated to help verify the predictive accuracy of V-3. Also, some participants indicated that students required more than three retestings to achieve the faculty-designated benchmark score, so future researchers might consider expanding the study to include investigating the predictive accuracy of more than three versions of the E^2.

Surveys of schools' remediation strategies have been conducted by numerous researchers. Although the information provided by these studies is interesting, the survey data do not describe the effectiveness of the various remediation strategies. Future researchers might consider conducting studies that implement an experimental design whereby the effectiveness of various remediation strategies can be compared.

CONCLUSION

The findings of S7 support the findings of the previous six studies in that the E^2 was again found to be a highly accurate predictor of NCLEX-RN success. The use of parallel versions of the E^2 to evaluate the effectiveness of remediation was also supported by the findings of this study, because all three versions of the E^2 had a predictive accuracy above 90%. Remediation seemed to be effective because additional students achieved scores of 900 and above with each additional retesting. However, further research is needed to compare the effectiveness of specific remediation strategies.

At a time when schools of nursing are battling budget cutbacks and faculty shortages, it is helpful to deans, directors, and faculty to have access to a measurement tool that can predict student success. Based on the findings of the seven studies that examined the validity of the E^2, schools can confidently use the E^2 to assess students' preparedness for the licensure examination.

REFERENCES

1. Lauchner K, Newman M, Britt R. Predicting licensure success with a computerized comprehensive nursing exam: the HESI Exit Exam. *Comput Nurs*. 1999;17(3):120–125.
2. Newman M, Britt R, Lauchner K. Predictive accuracy of the HESI Exit Exam: a follow-up study. *Comput Nurs*. 2000;18(3):132–136.
3. Nibert A, Young A. A third study on predicting NCLEX success with the HESI Exit Exam. *Comput Nurs*. 2001;19(4):172–178.
4. Nibert A, Young A, Adamson C. Predicting NCLEX success with the HESI Exit Exam: fourth annual validity study. *Comput Inform Nurs*. 2002;20(6):261–267.
5. Lewis C. *Predictive Accuracy of the HESI Exit Exam on NCLEX-RN Pass Rates and Effects of Progression Policies on Nursing Student Exit Exam Scores* [dissertation]. Houston, TX: Texas Woman's University; 2005.
6. Adamson C, Britt R. Repeat testing with the HESI Exit Exam—sixth validity study. *Comput Inform Nurs*. 2009;27(6):393–397.
7. Morrison S, Adamson C, Nibert A, Hsia S. HESI exams: an overview of reliability and validity. *Comput Inform Nurs*. 2004;22(4):220–226.
8. Nibert A, Young A, Britt R. The HESI Exit Exam: progression benchmark and remediation guide. *Nurse Educ*. 2003;28(3):141–145.
9. Crocker L, Algina J. *Introduction to Classical and Modern Test Theory*. Belmont, CA: Wadsworth-Thomson Learning; 1986.
10. Paul R. *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World*. Rohnert Park, CA: Center for Critical Thinking and Moral Critique; 1990.
11. Morrison S, Nibert A, Flick J. *Critical Thinking and Test Item Writing*. 2nd ed. Houston, TX: Health Education Systems Inc; 2006.
12. National Council of State Boards of Nursing. Alternate item formats frequently asked questions. https://www.ncsbn.org/2334.htm. Accessed July 28, 2011.

For more than 20 additional articles related to education, go to NursingCenter.com\CE.