Increasingly, Elsevier’s HESI Exit Exam (E²) is being used to assess students’ readiness for the National Council Licensure Examination for Registered Nurses (NCLEX-RN). Seven previously conducted validity studies indicate that the E² is 96.36%–99.16% accurate in predicting NCLEX-RN success. Findings of this eighth validity study, which also investigated the predictive accuracy of repeat testing with parallel versions of the E², indicated that the E² is highly accurate (94.93%–98.32%) in predicting NCLEX-RN success for the initial testing and 2 retests. Of the 66 participating nursing programs, deans and directors from 43 (65.15%) of the programs reported implementing a policy that used E² scores as a benchmark for remediation. A score of 850 was the most common E² benchmark designated by faculties, and students who failed to achieve the faculty-designated E² benchmark score were required to retest with a parallel version of the E². Remediation resources used to assist students in achieving faculty-designated E² benchmark scores varied widely, with many programs employing multiple remediation methods. (Index words: HESI Exit Exam; NCLEX-RN; Evaluation; Remediation; Testing policy) J Prof Nurs 29: S5–S9, 2013. © 2013 Elsevier Inc. All rights reserved.

The National Council Licensure Examination for Registered Nurses (NCLEX-RN) is a high-stakes exam for all involved. Graduates who fail the NCLEX-RN suffer a delay or loss of a promising career, as well as a sense of personal failure. Faculties frequently use the school’s NCLEX-RN pass rate as a measure of curricular effectiveness and may feel disappointed when program graduates fail the licensure examination. National accrediting agencies such as the Commission on Collegiate Nursing Education and the National League for Nursing Accrediting Commission (NLNAC), as well as state boards of nurse examiners, use NCLEX-RN pass rates as a benchmark for program effectiveness (American Association of Colleges of Nursing, 2009; National Council of State Boards of Nursing [NCSBN], 2009; NLNAC, 2008). An ongoing pattern of mediocre pass rates can place a nursing program’s accreditation at risk, and poor pass rates can hinder recruitment of well-qualified applicants and damage the college or university’s reputation. Therefore, the potential consequences of NCLEX-RN failure have long motivated nurse educators to search for measures of students’ readiness for the licensure examination. A number of studies indicate a moderate relationship between NCLEX-RN success and social factors—such as age, ethnicity, and socioeconomic status—whereas others describe NCLEX-RN success in relation to academic factors—such as Scholastic Assessment Test and American College Testing scores, prerequisite science course grades, senior-level nursing course grades, cumulative grade point average, and outcomes on a variety of teacher-made comprehensive exams (Arathuzik & Aber, 1998; Daley, Kirkpatrick, Frazier, Chung, & Moser, 2003; Endres, 1997; Giddens & Gloeckner, 2005; Haas, Nugent, & Rule, 2004; Higgins, 2005; Romeo, 2010; Schmidt, 2000; Yin & Burger, 2003).

Increasingly, schools of nursing are administering a commercially produced comprehensive exam such as Elsevier’s HESI Exit Exam (E²) as an indicator of students’ preparedness for the NCLEX-RN. The first published study regarding the use of the E² indicated that
54 registered nurse (RN) programs administered the E^2 during the academic year 1996–1997 (Lauchner, Newman, & Britt, 1999). By the 2007–2008 academic year, over 600 schools in 50 states had administered the E^2 (Elsevier database, August 2009). Since 1999, seven studies have been published that investigated the validity of the E^2, with samples that consisted of associate degree (ADN), baccalaureate degree (BSN), and diploma nursing programs throughout the United States (Adamson & Britt, 2009; Lauchner et al., 1999; Lewis, 2005; Newman, Britt, & Lauchner, 2000; Nibert & Young, 2001; Nibert, Young, & Adamson, 2002; Young & Willson, 2012). The total RN student sample size for these seven studies was 41,567, and findings indicated that the E^2 was 96.36%–99.16% accurate in predicting NCLEX-RN success. Each successive study built on and expanded the focus of the previous investigations. The sixth and seventh E^2 validity studies investigated the predictive accuracy of two retests with parallel versions of the E^2. These retests were administered to students who did not achieve the faculty-designated E^2 benchmark score on their previous testing. Findings of the sixth E^2 validity study indicated that there was no significant difference between the predictive accuracy of the first version (V1) and the second version (V2) of the E^2 but that the predictive accuracy of the third version (V3) was significantly lower than V1 and V2. However, findings of the seventh E^2 validity study indicated that the predictive accuracy of V1 was significantly higher than that of V2 and V3, but the range of accuracy for the all three versions was quite high (93.24%–99.16%). Because the predictive accuracy of V3 in the sixth validity study (82.50%) was more than 10% lower than the predictive accuracy of V3 for seventh validity study (93.24%), the authors of the seventh validity study recommended replicating the study to help verify the predictive accuracy of V3. Therefore, the purpose of this eighth E^2 validity study was to again examine the predictive accuracy of three parallel versions of the E^2. Additionally, testing protocols, remediation strategies, and progression policies implemented by the sample schools were explored.

**Methods**

The theoretical framework used to guide this study and the seven previously conducted E^2 validity studies was based on critical thinking theory described by Paul (1990) and classical test theory described by Crocker and Algina (2006). A detailed explanation of these theoretical frameworks and a conceptual model for exam development are described by Morrison, Adamson, Nibert, and Hsia (2004). Institutional review board approval was obtained prior to conducting this study. A descriptive, comparative, retrospective design was used to investigate the accuracy of three successively administered parallel versions of the E^2 in predicting NCLEX-RN success. Students’ E^2 scores were stored in Elsevier’s HESI database, and after obtaining NCLEX-RN outcomes from the deans and directors of the participating nursing programs, these data were also stored in the HESI database. No personal identification data, such as students’ names or the names of the participating schools were provided to the researchers.

**Instruments**

Two instruments were used to collect data for this study: the E^2 and the Eighth Validity Study Questionnaire. The E^2 is a 160-item, computerized, comprehensive exam that is designed to simulate the NCLEX-RN. Ten of the 160 test items are pilot items and are not considered in the calculation of the students’ scores but are included on the E^2 for the purpose of obtaining item analysis data. Difficulty level and discrimination data are obtained on every test item with each use. To be included in an E^2, each test item must have a difficulty level of no less than 40%, and the test items included in an exam must have an average discrimination index of at least 0.15. A proprietary mathematical model is used to calculate HESI scores, which range from 0 to over 1,000 and can be as high as 1,500. The highest score is dependent on the difficulty level of all the test items included in the exam. A Kuder–Richardson Formula 20 is calculated for all HESI exams administered, and these data are used to calculate the estimated reliability of an exam prior to its administration (Morrison et al., 2004). The E^2 estimated reliability for all exams administered to the study group was greater than 0.9. The E^2 distribution of content on all exams is the same as the distribution of content described in the most current NCLEX-RN blueprint (NCsBN, 2010), thus helping to ensure content validity of the E^2. Criterion-related validity is measured by evaluating the accuracy of the E^2 in predicting NCLEX-RN success, and the predictive accuracy of the last seven validity studies ranged from 96.36% to 99.16%, indicating that the E^2 is a valid exam (Adamson & Britt, 2009; Lauchner et al., 1999; Lewis, 2005; Newman et al., 2000; Nibert & Young, 2001; Nibert et al., 2002; Young & Willson, 2012).

The Eighth Validity Study Questionnaire was a 10-item electronic survey with multiple-response options that was designed to obtain data from participating schools about their policies regarding the use of E^2 scores as a benchmark for remediation, retesting, and progression. Deans and directors at participating schools were provided with a list of their students who had taken the E^2 between September 1, 2007, and August 31, 2008, and were asked to identify the students’ success or failure on their first attempt at taking the NCLEX-RN. After identifying students’ NCLEX-RN outcomes, respondents were told to remove all student names from the list before uploading the data to Elsevier. To ensure confidentiality, these data were stored in the HESI database as aggregates that included no student or school names.

**Sample**

Based on an a priori power analysis, a random sample of 154 nursing programs, stratified by program type, was selected from a population of more than 600 nursing programs that administered the E^2 between September 1,
2007, and August 31, 2008. The stratified random sample included 78 ADN, 69 BSN, and 7 diploma nursing programs. After the initial electronic mailing and two follow-up electronic mailings, deans and directors from 66 (42.86%) of the selected nursing programs responded to the survey: 36 (54.55%) ADN programs, 26 (39.39%) BSN programs, and 4 (6.06%) diploma programs. The student sample consisted of 3,962, of which 204 were removed from the sample because 109 students did not graduate and no NCLEX-RN outcome data were available for 95 students. Therefore, the total sample consisted of 3,758 nursing students: 2,159 (57.45%) ADN students, 1,500 (39.91%) BSN students, and 99 (2.63%) diploma students (see Table 1).

### Results

Those who score 900 and higher on the E2 are described by Elsevier, the publisher of the E2, as predicted to pass the NCLEX-RN. Of the 3,758 participants in this study, 2,685 took V1 only; and of the 1,132 who scored 900 and higher, 1,113 (98.32%) passed the NCLEX-RN on their first attempt. Students who did not achieve their school’s faculty-designated E2 benchmark score were often required to remediate and retest with a parallel version of the E2. Of the 3,758 participants, 881 students took both V1 and V2, and of the 355 who scored 900 and higher, 337 (94.93%) passed the NCLEX-RN on their first attempt. Of the 3,758 participants, 192 students took V1, V2, and V3, and of the 73 who scored 900 and higher, 70 (95.89%) passed the NCLEX-RN on their first attempt. The predictive accuracy was significantly different across all versions for those scoring 850–899 on one of the three versions of the E2, and of these, 1,520 (97.44%) passed the NCLEX-RN on their first attempt. Therefore, the accuracy of the E2 in predicting NCLEX-RN success was 97.44%, regardless of whether the student was required to take up to three versions of the E2 before achieving a score of 900 (see Table 2).

To examine E2 and benchmarking policies, deans and directors of the participating nursing programs were asked if they had established a policy that used E2 scores as a benchmark for remediation and, if so, to identify the faculty-designated E2 benchmark score and to describe the methods or strategies used to remediate students, the requirements placed on such remediation, and the consequences associated with failing to achieve that score. Furthermore, participants were asked to describe the methods or strategies used to remediate students, the requirements placed on such remediation, and the consequences associated with failing to complete the required remediation. Of the 66 participating nursing programs, deans and directors from 43 (65.15%) programs reported that they had established a remediation and retesting policy based on E2 scores. Of the 43 nursing programs that used E2 scores as a benchmark for remediation, 31 (72.09%) designated 850 as their E2 benchmark score, and 7 (16.28%) designated 900. The faculty at the remaining five schools designated a benchmark score other than 850 or 900. Because a score of 850 was so frequently cited as the faculty-designated benchmark score, the predictive accuracy of the E2 for those scoring 850–899 was examined. Of the 3,758 participants, 721 students scored 850–899 on one of the three versions of the E2, and 666 (92.37%) passed the NCLEX-RN on their first attempt. Findings indicated that of the 505 students who scored 850–899 on V1, 483 (95.64%) passed the NCLEX-RN on their first attempt; of the 189 students who scored 850–899 on V2, 167 (88.36%) passed the NCLEX-RN on their first attempt; and of the 27 who scored 850–899 on V3, 16 (59.26%) passed the NCLEX-RN on their first attempt. The predictive accuracy was significantly different across the three versions for those scoring 850–899, \( \chi^2 (2, N = 721) = 54.00, P < .01 \). A chi-square analysis indicated that the predictive accuracy was significantly different for those scoring 850–899 than it was for those scoring 900 and higher across all versions, \( \chi^2 (1, N = 2,281) = 31.68, P < .01 \). Furthermore, for those scoring 850–899, the

### Table 1. Description of Sample According to Program Type

<table>
<thead>
<tr>
<th>Program type</th>
<th>n</th>
<th>%</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADN</td>
<td>36</td>
<td>54.55</td>
<td>2,159</td>
<td>57.45</td>
</tr>
<tr>
<td>BSN</td>
<td>26</td>
<td>39.39</td>
<td>1,500</td>
<td>39.91</td>
</tr>
<tr>
<td>Diploma</td>
<td>4</td>
<td>6.06</td>
<td>99</td>
<td>2.63</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>100.00</td>
<td>3,758</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### Table 2. Significance and Predictive Accuracy Between Versions and Overall

<table>
<thead>
<tr>
<th>Scoring categories</th>
<th>V1 and V2</th>
<th>V1 and V3</th>
<th>V2 and V3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>900 and higher</td>
<td>1,132</td>
<td>1,132</td>
<td>355</td>
<td>1,560</td>
</tr>
<tr>
<td></td>
<td>98.32  *</td>
<td>98.32  *</td>
<td>94.93  *</td>
<td>97.44  *</td>
</tr>
<tr>
<td></td>
<td>12.82 *</td>
<td>2.26</td>
<td>0.12</td>
<td>13.18 *</td>
</tr>
<tr>
<td>850–899</td>
<td>355</td>
<td>73</td>
<td>73</td>
<td>721</td>
</tr>
<tr>
<td></td>
<td>95.64  *</td>
<td>95.89</td>
<td>95.89</td>
<td>92.37  *</td>
</tr>
<tr>
<td></td>
<td>12.29 *</td>
<td>58.32 *</td>
<td>15.46 *</td>
<td>54.00  *</td>
</tr>
<tr>
<td></td>
<td>88.36</td>
<td>59.26</td>
<td>59.26</td>
<td></td>
</tr>
</tbody>
</table>

Note. df = 1 for each comparison between versions and df = 2 for the overall comparisons.

* \( P \leq .01 \).
predictive accuracy was significantly different and lower with each successive E² taken (see Table 2).

Deans and directors at 38 participating schools identified the following consequences for failing to achieve the faculty-designated E² benchmark score: failure of a nursing course, 6 (15.79%); delayed or denied graduation, 17 (44.74%); and delayed or denied NCLEX-RN candidacy, 15 (39.47%) (see Table 3). Remediation was required by 47 (71.21%) of the 66 participating nursing programs. The duration of the required remediation for most of the schools (36 or 76.60%) ranged from 2 weeks to 6 weeks. Respondents indicated that a variety of strategies were used for remediation, including Elsevier’s online case studies and NCLEX-RN review books, as well as faculty-developed materials. Remediation methods ranged from face-to-face instruction to group tutoring and from online enrichment courses to repeating various nursing courses. Many schools used a combination of materials and methods.

**Discussion**

The findings of this study support those of the seven previously conducted E² validity studies. The predictive accuracy for the initial administration of the E² was 98.32% for the eighth validity study, which is within the range of the predictive accuracy for the previous seven studies, 96.36%–99.16%. However, the predictive accuracy of repeat testing with parallel versions of the E² differed among the three studies that investigated repeat testing: the sixth, seventh, and eighth validity studies. In the sixth validity study, the predictive accuracy of V3 was significantly lower than V1 and V2, but in the seventh validity study, the predictive accuracy of both V2 and V3 was significantly lower than V1. In this eighth study, there was a significant difference in the predictive accuracy between V1 and V2, but all three versions had a predictive accuracy higher than 94%, which indicates that the E² can be used with a high degree of confidence in its ability to predict NCLEX-RN success. It is worth noting that except for V3 in the sixth validity study, the predictive accuracy for all versions of the E² across all three studies was 92.94% or higher, and for this eighth study, the predictive accuracy for those scoring 900 and higher for all three versions was 97.44%. Based on these findings, faculty can be assured that the E² is highly accurate in predicting NCLEX-RN success, regardless of whether or not the student up to three testings to achieve a score of 900. Furthermore, it is likely that the predictive accuracy of V3 in the sixth validity study (82.50%) was a spurious finding because it is so different from the predictive accuracy of all E² versions evaluated in all three validity studies that examined repeat E² testing.

When interpreting the findings of this study, it must be noted that NCLEX-RN outcomes were not available for 204 students from the participating schools. It would be worthwhile to determine how many times they took the E² and what they scored. Because all identifying student information and nursing program data were deleted prior to uploading the data to the Elsevier database, it was not possible to examine these students to see if they differed from the participating sample. Despite the low number of students who did not take the NCLEX-RN, it might be worthwhile to follow these students longitudinally to determine if and when they took the NCLEX-RN and what their outcomes were on the licensure examination.

The seventh and eighth validity studies explored program practices including the use of E² scores for benchmarking, retesting, and remediation. The findings of both studies were similar. Most of the faculties designated 850 as their E² benchmark score, followed by 900. The findings of the seventh validity study indicated that the predictive accuracy was 96.86% for those who scored 850 and higher on the E², regardless of whether or not the students were required to take up to three versions of the E² (Young & Willson, 2012). This finding was based on a calculation that included not only those who scored 850–899 but also all those who scored 900 and higher. This eighth validity study determined that the NCLEX-RN pass rate for those scoring 850–899 was 95.64% for V1 but dropped to 88.36% for V2 and 59.26% for V3. This finding has implications for faculty when designating their E² benchmark score. If an E² score of 850 is chosen as the benchmark, students who score 850–899 need to be encouraged to remediate because their risk of failing the NCLEX-RN increases significantly with each repeat E² testing.

**Conclusions**

Based on the findings of this study, repeat testing with up to two parallel versions of the E² is highly predictive of NCLEX-RN success. Therefore, nursing faculties have evidence-based data to support their decision to establish testing protocols, remediation strategies, and progression policies that use E² scores as a benchmark. Faculties at most of the participating schools chose 850 as a benchmark E² score, which seems reasonable when reviewing the predictive accuracy for those scoring 850–899 on V1 of the E². However, faculties and students should be cautioned that those scoring 850–899 on repeat testing with the E²
have a significantly greater risk of failing the NCLEX-RN with each successive retesting. Therefore, continued remediation prior to taking the NCLEX-RN is essential for all students who score 850–899 on any version of the E², but it is of paramount concern for those who require repeat retesting to achieve a score of 850.

This eighth validity study confirms the finding of the previous seven validity studies—the E² is highly predictive of NCLEX-RN success. Therefore, nursing faculties can use the E² with confidence to assess students' readiness for the NCLEX-RN.

References


