Ahmadi Safa and Sheykholmoluki
Language Testing in Asia (2023) 13:40
https://doi.org/10.1186/s40468-023-00254-0

RESEARCH

Open Access

An impact study of the Iranian National University Entrance Exam from students and parents' perspectives

Mohammad Ahmadi Safa[1*] and Hamidreza Sheykholmoluki[1]

*Correspondence:
m.ahmadisafa@basu.ac.ir

[1] Department of English Language, Bu-Ali Sina University, Hamedan, Iran

Abstract

As a part of a larger project, this study reports a large-scale investigation of the impacts of the Iranian National University Entrance Examination (INUEE) on Iranian high school students and their parents. For this purpose, in a mixed methods research design, 1350 high school students selected based on convenience sampling from three western provinces of Iran were given a researcher-made Likert scale questionnaire; moreover, 27 parents sat for a semi-structured interview. The quantitative data obtained from the questionnaire were subject to inferential statistical analyses, and the qualitative interview data were first transcribed, and then the transcriptions were subjected to content analysis to extract common patterns and recurring themes. The integrated results obtained from the quantitative and qualitative data analyses indicated that from most of the participants' points of view, INUEE had detrimental impacts on students' and their parents' educational, personal, and social life. From students' perspective, INUEE had the most harmful impact on their physical and psychological well-being, and from parents' vantage point, the most negative impacts of INUEE were on the economy of the family and their social activities and entertainments. The findings of the study provide further evidence of the consequential invalidity of such large-scale high-stakes tests and the negative consequences following them. Furthermore, the attested negative impacts of high-stakes large-scale testing on the test takers, their family members, and the educational systems jeopardize educational justice in general and the usefulness and fairness of such tests in particular.

Keywords: Consequential validity, High-stakes test, INUEE, Test fairness

Introduction

High-stakes tests are defined as tests "whose results are seen rightly or wrongly by students, teachers, administrators, parents, or the general public as being used to make important decisions that immediately and directly affect them" (Madaus, 1988, p. 87). According to Moses and Nanna (2007), high-stakes tests are those "whose scores have a direct impact on a person's life options and opportunities" (p. 56). A test is high-stakes for test takers if, for instance, their chance of gaining admission to university depends on the test results and for teachers if their pay or promotion is affected by the test

outcomes, or if their reputation is determined by their students' test performance. A test can also be high-stakes for institutions if their reputation and/or funding are determined by their students' test scores.

High-stakes tests can not only have direct and strong effects on teaching and learning activities (Alderson & Wall, 1993), but they can also exert strong effects on teachers' and learners' attitudes, behaviors, and motivation (Shohamy, 1993). It is exactly on such basis that assessment systems are viewed as high-stakes when they have grave consequences for students (Thomas, 2005; Ysseldyke et al., 2004). Hence, the identifying factor for considering a test as high-stakes is the attribution of consequences to the outcomes. Such consequences have been increasingly at the focus of language testing researchers' attention ever since Messick's (1989) validity matrix introduction, a few years after which Alderson and Wall (1993) ironically ask "does washback exist?" and Brennan (2006) plainly claimed that perhaps the most controversial issue in validity has been the role of consequences.

The consequences incorporate the effects of test results (positive or negative; intended or unintended) on different stakeholders such as students, teachers, administrators, and larger societal systems (Cheng, 2008; Fulcher, 2014; Hubley & Zumbo, 2011; Kunnan, 2004; Shohamy, 2001). While researchers like Shohamy (2001) and Lynch (2001) consider the investigation of consequences as an integral part of validation studies, Bachman (2005) views the incorporation of consequences into validity research as an impractical endeavor and Davies (2008) assigns social and political accountability roles to such consequences, an approach that McNamara and Roever (2006) view with skepticism. On the other hand, according to Stecher and Barron (2001), "one important step that should be taken is to study the consequences of the testing system as rigorously as we study the reliability and validity of the scores" (p. 280). Madaus et al. (2009) favor an independent monitoring body to ensure the stakeholders that the tests are technically sound, negative consequences are minimized, and the payoffs outweigh the pitfalls.

Offering a delimited role to consequences in validity, Messick (1989; 1980) incorporates consequences into his validity matrix and argues that "scores are always connected to value implications, which make a basis for score interpretation and use and connect construct validity, consequences, and policy decisions" (Chalhoub-Deville, 2009, p.120). Kane (2006) and McNamara (2006) praise Messick (1989) for elucidating the role of consequences in validity theory. However, Kane (2013) argues that consequential research at the level of score interpretation, although necessary, is not sufficient to confirm the quality of test outcomes. According to Kane, even when score interpretation is sensible, poor decisions can be made. He believes "the evaluation of test score uses requires an evaluation of the consequences of the proposed uses, and negative consequences can render a score use unacceptable" (Kane, 2013, p. 46). While Kane (2013) does not elaborate on how research should be carried out to deal with construct-related consequences, through his collaborative work on the theory of action (Bennett et al., 2011), validity research finds the capacity to add a social dimension to consequences.

Theory of action (TOA) addresses consequential research at various levels, from individual scores to aggregate results, including teachers, administrators, and schools, as well as the social and educational contexts of tests. As Bennett (2010) put it, TOA

research "gives greater prominence to the effects of the assessment system on individuals and institutions as well as to the underlying mechanisms behind those effects" (p. 71). TOA demands a social perspective of research into consequences to deal with not only intended but also unintended effects. Through TOA research, data is collected "from key stakeholders (students, parents, teachers, and administrators) documenting how assessment results are used, noting both intended and unintended consequences of score use" (Bennett et al., 2011, p. 4).

Studies documenting the negative consequences of high-stakes tests are not scarce. One of the oft-mentioned negative consequences of high-stakes testing, for instance, is the stress and anxiety experienced not only by students, but also by all stakeholders. Research has shown that high-stakes tests can cause many students to experience a debilitating level of emotional and physical stress (Watson et al., 2014). Furthermore, teachers, administrators, and parents have all voiced concerns regarding the stress and anxiety involved in high-stakes testing (Amrein & Berliner, 2003; Barksdale & Thomas, 2000; Landry, 2006).

Another negative consequence of high-stakes testing is that such tests can kill students' thinking abilities (Fitzgerald, 2015). Furthermore, as a consequence of such tests, the type of instructional activities implemented in classrooms have changed from cognitively demanding tasks involving higher-level thinking to a focus on rote practices specifically intended to teach students the tips and tricks of multiple-choice tests (Ahmadi Safa & Sheykholmoluki, 2023; Barksdale & Thomas, 2000; Dong, 2020; Guloglu-Demir, & Kaplan-Keleş, 2021; Rezaeian et al. 2020; Sheperd, 2003; Tsang & Isaacs, 2022). Ysseldyke et al. (2004) maintain that "some researchers empathetically believe that students are failing to develop higher-order thinking skills as a result of drill-and-practices of teaching methods that hope to improve student performance in high-stakes exams" (p. 84). Brady (2008, p.66) points out that "of all the obstacles to improving student thinking, these [high-stakes tests] are surely the most damaging." Walker (2014) voices concern over how classroom educators prepare future test takers rather than creative and critical thinkers as a result of the consequences of high-stakes testing.

The negative social impacts of high-stakes testing have also been reported in many studies. According to Valenzuela (2000), such tests objectify students and negate them as individuals and as cultural beings with distinct experiences, needs, and desires that accompany their differences. Ahmadi Safa and Sheykholmoluki (2023) showed that high-stakes tests have negative impacts on students' friendships as well as their relationships with their family members, teachers, principals, and even school staff. Booher-Jennings (2008) also noticed that students established a moral hierarchy as a result of their high-stakes test performance. That is, some students who had a better performance on the test did not have a desire to associate with the students who had difficulty. Even so, there have been reports of bullying being linked to high-stakes testing. Hazel (2010), for example, reported that there was an interaction between high-stakes testing and student bullying. It was also shown that there was a direct relationship between teachers' stress and student bullying. That is, as the teachers' stress went up, student bullying also increased.

Reports of even physical illnesses as a result of high-stakes testing are not rare either. As Watson et al. (2014) indicated "the range of symptoms included various types of

physical pain, exhaustion, and nausea" (p. 6). Many students in Watson et al. (2014) said that they were exhausted and sleepy before the test even started. They had not been able to sleep the night before the test because of the anxiety and pressure of the test, and they felt frightened and powerless.

On the basis of the brief literature reviewed above, it is quite evident that the negative consequences of high-stakes tests have been a real concern for language testing experts and researchers. Even Bracey (2003) holds that it is not far-fetched to assume that a high-stakes test can cruelly and criminally function as a weapon of mass destruction in that it can have detrimental and far-reaching consequences on test-takers and all other stakeholders. On the basis of such strong and alarming reasoning, it is quite logical to assume that such negative consequences might be conceivable for INUEE, which serves as a nationwide high-stakes selection test; however, such negative consequences have been only scantily documented and quite limited in scope and number. Prior to a brief review of a number of studies focusing on INUEE, a brief description of the Iranian general education system is in order. The Iranian educational system is divided into K-12 education and higher education. The K-12 education, which is carried out by the Ministry of Education over a period of 13 years, includes a single year of pre-primary kindergarten (optional), 6 years of primary school, 3 years of junior high school, and 3 years of senior high school. It is at this last level that students choose their majors and typically spend most of their time preparing for INUEE. At the end of this period, students partake in nationwide final examinations to receive a high school diploma which is a requirement to take part in INUEE and possibly acquire admission into higher education centers and universities.

The few studies which have explored INUEE consequences have confirmed the potentiality of INUEE to seriously affect Iranian high school students' cognitive, affective, and social aspects. Ghorbani (2008) for instance explored the nature and scope of the impacts of INUEE on high school English teachers and verified that, regardless of their experience, educational background, and gender, the teachers confirmed the negative effects of INUEE on their curricular planning and instruction. Salehi and Yunus (2012) also explored high school English teachers' perceptions towards the INUEE and reported negative perceptions of the participating teachers towards the test. Finally, Ajideh and Mahmoudi (2017) investigated the washback effects of the English section of INUEE and concluded that the students' perceptions towards INUEE were a mixture of both positive and negative ones. Against this backdrop, this study, as a rather large-scale project only a part of which is reported in this paper, addressed different groups of stakeholders including students, their parents, high school teachers, and high school principals' attitudes towards INUEE impacts; however, due to space limitations, the current study reports only the results gained from Iranian high school students and their parents. For the stated purposes and drawing on both quantitative and qualitative data, the present mixed methods study aimed to explore the nature and extent of the impacts of INUEE on Iranian high school students and their parents, and the following research questions were raised:

> RQ1: What is Iranian high school students' opinion about the impacts of INUEE on different aspects of their life?

RQ2: What is the opinion of INUEE test takers' parents about the impacts of INUEE on different aspects of their family life?

Method

Design

The purpose of this study was to investigate the different educational, social, economic, and familial impacts of INUEE on the test takers from high school students and their parents' points of view. To this aim, the researchers applied a sequential exploratory mixed methods research design to collect quantitative and qualitative data, analyze them through appropriate procedures, integrate the results, and finally interpret the findings. The quantitative survey-type data were gathered through a researcher-made Likert scale questionnaire and subjected to descriptive and inferential statistical analyses. The qualitative semi-structured interview data on the other hand were gathered and analyzed following the Grounded Theory (Glaser & Strauss, 1967).

Participants

A total of 1350 high school students (745 males and 605 females) participated in this study. The participants were selected from three western provinces of Iran, i.e., Hamedan, Kermanshah, and Kurdistan, based on convenience sampling, in such a way that from each province, three of the most populated cities were selected. Next, 150 high school students were conveniently selected from each city, yielding a total of 1350 students. In sum, 893 (66.14%) of the participants were third grade, and 457 (33.85%) were second grade high school students. Of the 1350 students, 472 (34.96%) were majoring in humanities, 487 (36.07%) were majoring in experimental sciences, and 391 (28.96%) were majoring in mathematics. The second group of participants included a sample of 27 parents. For the selection of parents for the interview phase, the same sampling strategy was followed in a way that from each city, 3 parents (either the students' father or mother) who were available and willing to sit for an interview were selected, yielding a total number of 27 parents (16 males and 11 females). The parents ranged in age from 39 to 54. Out of the 27 parents, 5 did not have any educational degree (18.51%), 8 were high school graduates (29.62%), 9 were B.A. holders (33.33%), 4 M.A. holders (14.81%), and a single parent was a Ph.D. holder (3.70%). It should be noted that permissions for administering the questionnaires and the interview with students' parents were obtained first from the local offices of the Ministry of Education of the districts by completing the request forms and then from students and their parents. Being aware of the significance of the INUEE for the Iranian education system, students and their parents expressed their co eagerness to participate in the study, and their informed consent was thus obtained. The students and their parents were assured that their participation was on a voluntary basis and that the data obtained would be kept confidential.

Table 1 Reliability estimate of the questionnaire

Cronbach's alpha	Cronbach's alpha based on standardized items	No. of items
.941	.940	60

Table 2 The result of the KMO and Bartlett's test for the questionnaire

Kaiser-Meyer-Olkin measure of sampling adequacy		.716
Bartlett's test of sphericity	Approx. χ^2	2488.685
	df	1225
	Sig.	.000

Instruments

The instruments applied in this study included a researcher-made Likert scale questionnaire addressing the impacts of INUEE and a semi-structured interview designed to elicit the parents' opinions about the impacts of INUEE.

INUEE Impact Questionnaire

Motivated by a lack of any comprehensive and relevant questionnaire, the researchers set out to develop a questionnaire on the impact of the INUEE high-stakes test on Iranian high school students. To develop the questionnaire, at first, a pool of 100 high school students' opinions about INUEE were collected in written form. The students were asked to raise any comments or topics they felt were relevant to the impacts of INUEE. Then, drawing on the most recurring themes of students' opinions and the theoretical principles of test impact studies, a Likert-scale questionnaire consisting of 60 5-point Likert items on a scale of 1 (strongly disagree) to 5 (strongly agree) was developed. Next, to ensure the construct validity of the questionnaire, it was subjected to a number of testing and assessment experts' judgments. The experts commented on the accuracy and adequacy of the items. Having refined the problematic aspects of the scale on the basis of the experts' judgment, the researchers pilot-tested the scale with 200 high school students. To assess the reliability of the questionnaire, Cronbach's alpha measure of internal consistency was run on the pilot administration data. The results confirmed a high level of internal consistency for the scale ($\alpha = .94$), as shown in Table 1.

In addition, in an attempt to statistically validate the scale, principal component analysis (PCA) with Varimax rotation was carried out on pilot data. The initial results of the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) turned out to be 0.54 which was not considered as a highly acceptable level of sampling adequacy. On this basis, and in an attempt to improve the sampling adequacy, ten questionnaire items with low factor loading levels (below 0.40) were eliminated from the scale, and a second PCA was carried out on the curtailed data, and the resulting Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's Test of Sphericity were found acceptable (Table 2).

As is evident in Table 2 above, the KMO measure of sampling adequacy was 0.71, above the recommended value of 0.60 (Hair et al., 2010; Pallant, 2000; Tabachnick &

Fidell, 2007), and Bartlett's Test of Sphericity was statistically significant (X^2 (1225) = 2488.685, $P < 0.05$). Once the suitability of data for factor extraction and retention was verified, a meticulous inspection of the correlation matrix showed the presence of a large number of coefficients of 0.3 and above. PCA extracted 13 components with eigenvalues above 1, accounting for 53.38% of the variance. However, an inspection of the screeplot revealed a clear break (an elbow) at the seventh component. Thus, based on the screeplot elbow and the interpretability of the factor solution, it was decided to retain the first seven components with the highest factor loadings and the maximum amount of variance for further investigation. On this basis, a total of 18 items in the two groups were decided to be made redundant from the final scale: (1) the items which loaded on more than one component and (2) the items which minimally loaded on components 8 to 13 (mostly one item per component). After removing these items, it was deemed necessary to rerun the Cronbach's alpha measure of internal consistency and the KMO measure of sampling adequacy on the 32-item questionnaire; the Cronbach's alpha measure of internal consistency had slightly increased to .95 and the KMO measure of sampling adequacy also jumped to .89. The resulting seven-factor solution explained a total of 48.19% of the variance, with components 1 to 7 contributing 21.20, 5.69, 5.46, 4.78, 3.84, 3.78, and 3.44% of the variance, respectively. Upon the examination of the rotated factor matrix, it became clear that the first component, consisting of 8 items, was related to *students' well-being*; the second component including 7 items, was associated with *students' familial relationships*; the third component comprising 4 items was associated with *educational impact of INUEE*; the fourth component, including 4 items, was related to *the impact of INUEE on students' social life*; the fifth component consisting of 4 items was about *the economic impact of INUEE*; the sixth component consisting of 2 items was about *INUEE impacts on students' entertainments*; and the seventh component including 3 items was related to *negative impact of INUEE on students' motivation and self-confidence.* The final questionnaire comprising 32 items was thus designed and validated to explore the impacts of INUEE on different aspects of high school students' life (Additional file 1: Appendix A).

Semi-structured interview

In addition to the questionnaire and in line with the objectives of the study, in order to elicit the parents' attitudes about the impacts of INUEE on individual and social aspects of the test takers' life, a set of interview question items were formulated considering the theoretical principles of test impact studies with an eye to the factor structure of the INUEE Impact Questionnaire. In order to be able to ask follow-up questions and to delve deeper into the given matter, we chose to employ a semi-structure interview (Macqueen et al., 2018). Interviews were carried out either face-to-face or on the phone, and each lasted for nearly 30 min. Face-to-face interviews were conducted in a class either at school or at one of the INUEE institutes of the districts. The researchers' choice to conduct the face-to-face interviews in one of the classes in schools or institutes of the districts was because it was more convenient for the participating parents. The interview questions were so designed to explore various impacts INUEE had on students. Yet, the participants were asked to raise any topics they felt were relevant to the research

focus. To avoid any communication problems (Mackey & Gass, 2005), all the interviews were done in the participants' first language, i.e., Persian. The mother tongue was used in interviews so that the participants could freely express themselves and talk about their opinion easily with no language-related barrier. All the interviews were recorded for subsequent transcription, content analysis, and translation (Macqueen et al., 2018). To ensure the validity of the interview questions, two testing and assessment experts viewed and judged the validity of the items for the stated purposes (Additional file 1: Appendix B).

Data collection procedure

As described in the Instruments section, the INUEE Impact Questionnaire was developed, pilot tested, and validated on the basis of expert judgment and factorial analyses of the pilot data. Prior to the administration of the designed and validated INUEE Impact Questionnaire to the study sample, ethical principles of research data collection were considered. Hence, at first, permission for the distribution of the questionnaire was obtained from the offices of the Ministry of Education of the districts. Next, permissions were obtained from the intended high school teachers and principals. Being cognizant of the significance of the INUEE for the Iranian general education system, the school teachers and principals all graciously agreed to cooperate in the distribution of the questionnaire and the collection of data. Moreover, the participating students also showed their enthusiasm for completing the questionnaire and expressed their informed consent to participate in the project. However, in order to guarantee the response validity of the data, the participants were re-informed that their participation was only on a voluntary basis, and there was no obligation to take the questionnaire if anyone did not like to take it. Following such preliminary steps, the INUEE Impact Questionnaire was administered to a total of 1350 Iranian high school students selected on a convenience sampling basis by one of the researchers, and the required data was collected over a period of 3 months. In addition to the data collected from the high school students, the semi-structured interview described above was held with the second group of participants, i.e., a sample of high school students' parents, to elicit their opinions about different aspects of the INUEE impact. Each interview was carried out either face-to-face or on phone calls and lasted for about 30 min. All the interviews were audio-recorded.

Data analysis

At the initial data screening step, out of 1350 collected questionnaires, 315 were excluded from the analysis because they had been either left incomplete or were haphazardly taken by marking the same answer for all the questionnaire items. The remaining 1035 completed questionnaires were subject to descriptive statistical analyses. With respect to the qualitative data obtained from the interviews with parents, they were transcribed and subsequently subjected to content analysis to extract common patterns and recurring themes. After different levels of coding and "quantitizing" the data (Dornyei, 2007), the basic themes were identified, and their frequency of occurrence was counted and tabulated.

Results

INUEE Impact Questionnaire results

As already mentioned above, the INUEE Impact Questionnaire comprised seven factors labeled as *Impacts of INUEE on students' well-being*, *Impacts of INUEE on students' familial relationships*, *Educational Impacts of INUEE*, *INUEE Impacts of on students' social relationships*, *Economic Impacts of INUEE*, *INUEE impacts on students' entertainments*, and *Impacts of INUEE on students' motivation and self-confidence*. In order to answer the first research question of the study, i.e., What is Iranian high school students' opinion about the impacts of INUEE on different aspects of their life?, the data obtained for each one of the abovementioned components were extracted and tabulated separately and are presented one by one as follows.

Component 1: Impacts of INUEE on students' well-being

High school students' opinions about the impact of INUEE on their physical and psychological well-being were explored through the first component including eight items. Table 3 presents the results for this factor. It is noteworthy that due to space limitations within the tables, only the keywords of each questionnaire item are provided in Table 3 and all subsequent similar tables.

As shown in Table 3, the means for the first seven items ranged from 3.03 to 4.38, showing that the items enjoyed moderate to high means and only the last item (item 30) had a low mean score. Among the eight items of the first component, three items (i.e., items 7, 9, and 11) were related to students' psychological well-being, and five items (i.e., items 16, 17, 18, 19, and 30) were related to their physical well-being. Regarding the impact of INUEE on students' psychological well-being, the item that received the highest mean was addressing the anxiety that students experience as a consequence of INUEE (item 7), and the item with the comparatively lowest mean was associated with the symptoms of psychological diseases (item 9). With regard to the impact of INUEE on students' physical well-being, the items receiving the highest means were related to the negative impact of INUEE on students' diet (item 17) as well as the symptoms of physical diseases (item 18), while the item with the lowest mean was associated with using drugs for getting good results on INUEE (item 30). As the descriptive statistic data in Table 3 indicate, the negative impact of INUEE on the psychological and physical well-being of high school students is verified and

Table 3 Descriptive statistics for students' opinions about the impact of INUEE on their well-being

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
7. Anxiety of INUEE	4.38	1.01	3.3	4.3	7.2	21.6	63.6
9. Psychological diseases	3.58	1.30	11.2	12.8	8.8	41.1	26.1
11. Seeing a psychiatrist/psychologist	4.01	1.10	4.3	4.8	19.5	28.1	43.3
16. Physical damages caused by INUEE	3.03	1.38	19.2	17.4	24.0	20.4	19.0
17. Disorganized diet	4.05	1.19	5.3	10.5	5.3	31.6	47.3
18. Physical diseases	3.98	.98	3.8	5.2	10.5	50.5	30.0
19. Experience of medical treatments	3.04	1.44	21.4	17.1	18.7	21.4	21.4
30. Using drugs	2.97	1.35	18.9	20.3	21.3	23.4	16.1

Table 4 Descriptive statistics for students' opinions about the impact of INUEE on their familial relationships

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
5. Rift in the family	3.96	1.15	6.6	7.9	4.6	44.4	36.5
6. Argument with family members	3.74	1.07	5.9	9.9	9.2	54.6	20.4
8. Parents' unpleasant comparisons	3.53	1.44	13.5	15.7	10.4	25.4	35.0
12. Parents' frightening threats about failing INUEE	3.53	0.94	5.2	5.3	31.6	47.4	10.5
13. Fear of being excluded by parents	2.83	1.12	34.8	11.0	19.0	30.5	4.7
15. Parents' intolerable scolding	3.05	1.42	20.2	18.5	18.7	21.8	20.8
31. Negative familial relationships	3.57	1.03	4.8	10.0	25.2	43.8	16.2

Table 5 Descriptive statistics for students' opinions about the educational impact of INUEE

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
24. Negative impact on educational goals	3.88	.94	2.4	7.1	15.2	51.0	24.3
25. It leads to the discovery of students' talents	3.79	1.34	9.6	9.5	16.4	20.7	43.8
26. Negative impact on learning life skills	4.36	.76	1.0	1.4	7.6	41.0	49.0
28. Inability to measure students' achievements	3.89	1.30	8.8	8.1	13.8	24.4	44.9

documented. To the dismay of the test organizers and designers, more than 85% of the respondents agreed that the test brings about a high level of anxiety for them, and more than 80% reported that they were suffering from physical diseases as a result of this high-stakes test.

Component 2: Impact of INUEE on students' familial relationships

The second component of the questionnaire includes seven question items taped into the students' opinions about the impact of INUEE on their familial relationships. Table 4 summarizes the descriptive results for this factor.

As demonstrated in Table 4, the means of the items ranged from 2.83 to 3.96, showing that except for item 13, all other items had a moderate to high means which indicates that students so agreed with the items associated with the negative impact of INUEE on their familial relationships.

The items with the highest mean scores for this component were related to the rift in the families created as a consequence of INUEE (item 5) as well as the students' arguments with family members over INUEE (item 6), while the items receiving the lowest means were associated with students' fear of being excluded by their parents if they do not get good results on INUEE (item 13) and also their parents' intolerable scolding about their performance in INUEE (item 15). According to obtained results, it is descriptively verified that the familial relationships of the potential test takers and their family members have been at least moderately affected by INUEE. As exemplar evidence in this regard, 81% of the respondents either agreed or strongly agreed that a rift in the family is only one of the consequences of the test in their familial relationships.

Table 6 Descriptive statistics for the impact of INUEE on students' social relationships

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
1. Negative relationship with relatives	3.57	1.20	6.9	13.5	21.2	32.1	26.3
2. Negative relationship with teachers	3.87	1.31	8.7	9.4	13.4	23.5	45.0
3. Negative relationship with classmates	3.79	1.16	11.9	37.1	19.5	23.3	8.2
4. Negative relationship with other people	4.21	.73	1.1	2.8	10.9	47.1	38.1

Table 7 Descriptive statistics for students' opinions about the economic impact of INUEE

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
20. Financial pressure of buying INUEE materials	3.43	1.28	9.0	13.3	19.0	28.7	30.0
21. Financial pressure of enrolling in INUEE classes	3.48	1.33	13.2	9.5	20.9	28.6	27.8
22. Having to work for earning INUEE costs	3.38	1.54	19.8	12.6	12.6	19.5	35.5
23. Embarrassment of asking parents for more money to pay for INUEE costs	4.15	.81	4.0	4.3	9.5	35.2	47.0

Component 3: Educational impact of INUEE

The educational impacts of INUEE were investigated through the third component including 4 items. Table 5 presents the results for this factor.

As Table 5 demonstrates, the means of the items ranged from 3.79 to 4.36. As is evident below, the item with the highest mean score was related to the negative impact of INUEE on learning crucial life skills like creative and critical thinking, decision-making ability, and management of emotions (item 26), while the item that received the lowest mean score was associated with the incapability of the test for the identification of the students' talents and abilities (item 25).

Component 4: Impact of INUEE on students' social relationships

The fourth factor of the questionnaire comprises four question items taped into the students' opinions about the impact of INUEE on their social relationships. Table 6 summarizes the descriptive results for this factor.

As indicated in Table 6, the means of the items ranged from 3.57 to 4.21, showing that the majority of the respondents were of the opinion that INUEE had negative impacts on their social relationships. The item with the highest mean score was related to the negative impact of INUEE on students' social relationships with other people (item 4). On the other hand, the item receiving the lowest mean score was related to the negative impact of INUEE on students' relationships with their relatives (item 1). It is quite disappointing that 68% agreed or strongly agreed that the test leads to their negatively affected relations with their teachers.

Component 5: Economic impact of INUEE

The fifth factor of the questionnaire consists of four question items reflecting the students' opinions about the economic impact of INUEE. The descriptive results gained for this factor are summarized in Table 7.

Table 8 Descriptive statistics for students' opinions about the INUEE impacts on entertainments

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
29. No time for entertainment and recreational activities	4.18	.80	.5	3.3	12.4	45.7	38.1
32. No time for sports	3.70	1.32	9.2	12.1	15.6	25.3	37.8

Table 9 Descriptive statistics for students' opinions about the impact of INUEE on students' motivation and self-confidence

Item	Mean	SD	St disagree	Disagree	Neutral	Agree	St agree
10. Feeling a lack of self-confidence about performance in INUEE	3.84	1.36	9.3	12.1	10.1	22.3	46.2
14. Feeling despair about performance in INUEE	3.86	1.01	3.8	6.7	16.2	46.7	26.6
27. Feeling that efforts to get a high score in INUEE are ineffective	3.27	1.14	4.4	13.8	15.2	37.6	29.0

As shown in the table, the means of the items ranged from 3.38 to 4.15, meaning that the respondents relatively believed in the negative impact of INUEE on their economy. The item that students agreed with most in this component was associated with their embarrassment to ask their parents for more money to pay for INUEE costs (item 23), while the item that students agreed with least was related to having to work for earning INUEE costs (item 22).

Component 6: INUEE impacts on entertainments

High school students' opinions about the impact of INUEE on their entertainments were explored through the sixth component including 2 items. Table 8 presents the results for this factor.

As Table 8 shows, the mean score for the item related to students' opinions about the negative impact of INUEE on playing sports (item 32) was relatively high; however, the negative impact of INUEE on entertainment and recreational activities including outgoing and trip (item 29) received an even higher mean score.

Component 7: Impact of INUEE on students' motivation and self-confidence

Finally, high school students' opinions about the impact of INUEE on their motivation and self-confidence were investigated through the seventh component including 3 items. Table 9 presents the results for this factor.

As shown in Table 9, the mean scores for all three items were above average which denotes that the respondents mostly agreed that INUEE had negative impacts on their motivation and self-confidence. The item with the highest mean score in this component was related to students' opinions about having a feeling of despair about their performance in INUEE (item 14), while the item receiving the lowest means was

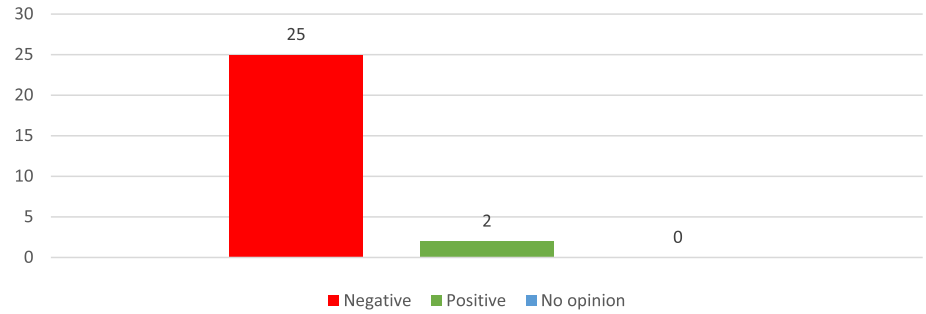

Fig. 1 Parents' feelings about INUEE

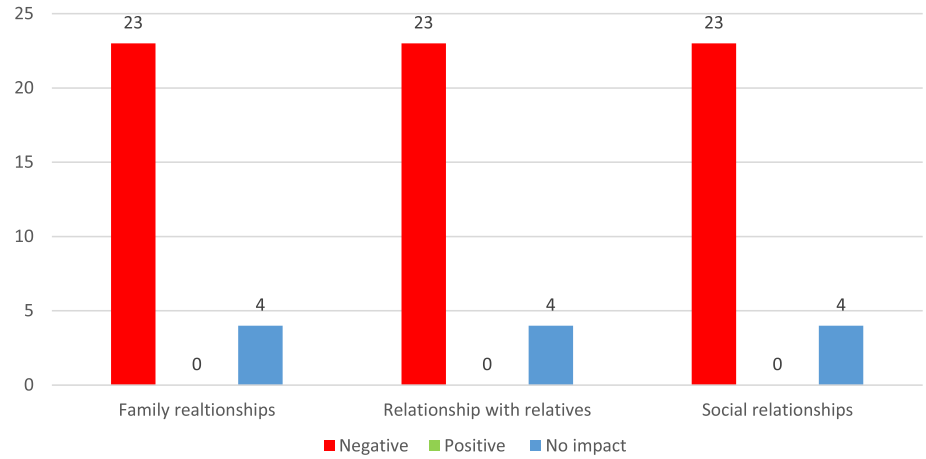

Fig. 2 Parents' opinions about the impacts of INUEE on relationships with others

associated with students' feeling that their efforts for getting a high score on INUEE will be ineffective (item 27).

Interview results

As regards the second research question concerning parents' opinions about the impacts of INUEE, parents' interview transcripts were subjected to content analysis, and the common patterns and recurring themes were extracted, codified, and quantitized. In addition to a brief description, the results are graphically displayed in bar graphs in the following part.

To begin with, parents were asked to express their feelings about INUEE. As is demonstrated in Fig. 1, 25 (92.59%) parents had negative feelings about INUEE, and only 2 (7.40%) were rather positive about it. Most parents described it as an "obstacle for success," "a dam for progress," "a grave for talents," "a nightmare," and even "a monster."

The parents were also asked to express their opinions about how INUEE impacted family relationships with others. As Fig. 2 demonstrates, out of 27 parents interviewed, 23 (85.18%) argued that INUEE had negative impacts on their familial and social relationships, and relations with their relatives, and only 4 (14.81%) parents

believed that the test has no impact on their relationships. With regard to the impact of INUEE on familial relationships, the majority of parents complained that preparing for INUEE has remarkably reduced their family gatherings. Also, some parents expressed that arguments over INUEE has created tension and at times even an atmosphere of hostility between family members. As one parent expressed, "there's always argument among my children even over trivial matters such as keeping the TV on or off." Another parent said, "my daughters used to take turn doing household chores. But, now that one of them has to take INUEE, they're always having quarrels over who does what in the house." One parent even stated, "our daughter is always falling out with us over why we spend too much money on buying INUEE materials and INUEE classes for his brother." As for the impact of INUEE on the relationships with relatives, most of the parents argued that visiting relatives is reduced as a result of their children's preparation for INUEE. One mother, for example, argued that, "my son doesn't like to have guests over to our house because he needs a quiet environment for studying for INUEE, and if one of the relatives drops in, he starts nagging me." Another parent said, "one of our close relatives has cut ties with us because when they wanted to pay a visit to us one night, we declined their request so that it doesn't interfere with our son's study time." Regarding the impact of INUEE on social relationships, most parents expressed that their children have become almost isolated and asocial since their children started preparing for INUEE, and this issue has in turn affected the social relationships of the whole family. Some of them argued that their children used to hang out with friends, go to the gym, etc., but all such activities are stopped since they have started preparing for INUEE.

As for the third interview question, the parents were asked about the impacts of INUEE on the economy of their family and also on family entertainments; the results of which are summarized in Fig. 3.

As Fig. 3 indicates, all parents (100%) argued that INUEE had negative impacts on both their family economy and family entertainments like going to a park, playing games, and trips. The majority of the parents expressed that the costs associated with INUEE materials and test preparation classes were way beyond their monthly income, especially in

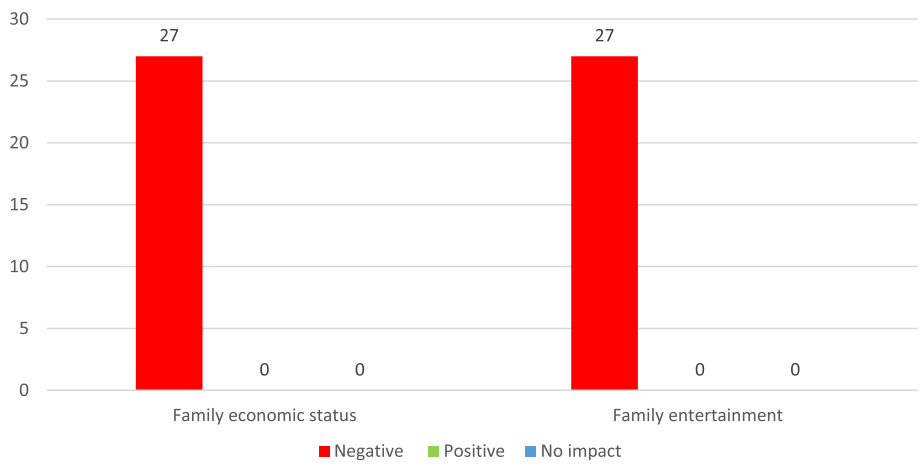

Fig. 3 Parents' opinions about the impacts of INUEE on family economic status and recreational activities

recent years in which the costs of INUEE materials and classes have skyrocketed. One parent argued, "having to pay for INUEE materials and classes has put the family under economic pressure; we have been even compelled to take out loans to be able to afford such costs." Another parent expressed, "since our boy has started taking INUEE classes and buying INUEE materials, we have been forced to stick to a tight budget and have not been able to save up money for even a computer." As for the impact of INUEE on family entertainments, most parents complained that having to prepare for INUEE has taken time away from other recreational activities. One parent stated that, "we used to go out every weekend, but now our whole life has become preparing for INUEE; this is depressing!" Another parent mentioned, "we can't even watch our favorite TV program. I wish there were no INUEE. We will surely breathe a sigh of relief after that."

The next question the parents were asked was about their opinions concerning the impacts of INUEE on the physical and psychological well-being of the whole family; the results of which are depicted in Fig. 4.

As Fig. 4 denotes, 13 (48.14%) parents argued that INUEE had negative impacts on their family's physical well-being, while 14 (51.85%) of them expressed that INUEE had no impact on their family's physical well-being. Among the parents who expressed that INUEE had negative impacts on their children's physical well-being, some referred to physical problems like exhaustion, severe headaches, hair loss, and hair cyst as a consequence of sedentary round-the-clock preparation for INUEE and even having the experience of being hospitalized in coronary care unit (CCU) because of INUEE trauma. One of the parents said "my daughter has suffered from severe stomachache since she began studying for INUEE. One night while studying she was writhing in such pain that we had to take her to hospital. We're always telling her to reduce her study time. I'm afraid INUEE will cause more serious health problems for her." Besides, 25 (92.59%) parents expressed that INUEE had negative psychological impacts on their children, and only 2 (7.40%) parents expressed that it had no impact at all. Among the psychological problems, the majority of the parents (88.88%) referred to stress and anxiety. One parent stated, "my daughter is always saying what happens if I fail INUEE, which gives her a lot of anxiety. Even, she sometimes has anxiety attacks which make her lose her appetite. Anxiety has also caused her to suffer from insomnia." Another parent said, "My son is constantly worried about his exam. He's always competing so hard with his classmates

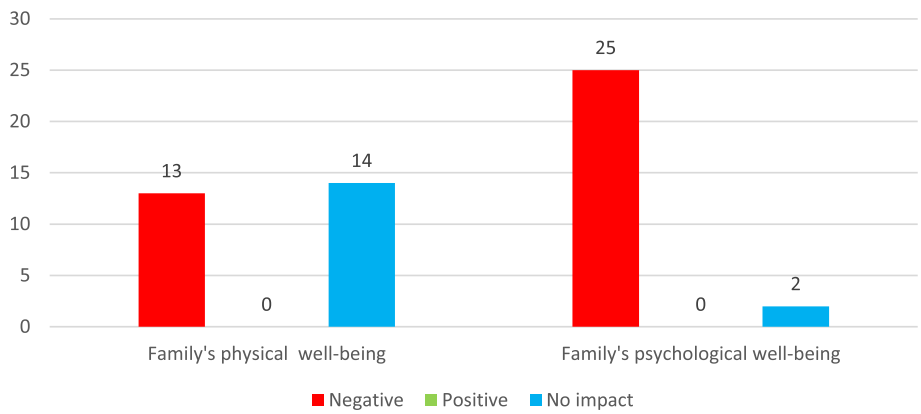

Fig. 4 Parents' opinions about the impacts of INUEE on physical and psychological well-being of the family

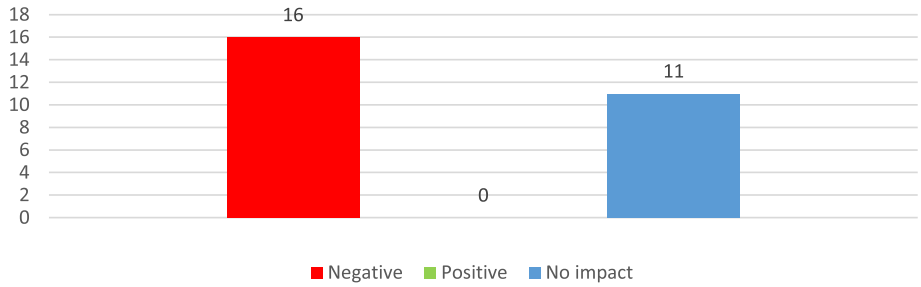

Fig. 5 Parents' opinions about the impacts of INUEE on students' upbringing

which makes him really anxious and his anxiety is injected to us as well. We are really worried about his future." Parents also referred to other psychological problems such as overaggressive behaviors, demotivation, depression, fear, and lack of self-confidence.

Parents were also asked to express their opinions about the impacts of INUEE on their children's upbringing; the results of which are summarized in Fig. 5.

As Fig. 5 shows, 16 (59.25%) parents complained that INUEE had negative impacts on their children's upbringing, while 11 (40.74%) parents expressed that INUEE had no impact on it. Among the parents who believed in the negative impact of INUEE in this regard, most of them argued that over-attention to getting an acceptable score in INUEE had unintentionally caused them to neglect to do their parenting responsibilities in bringing up their children well during the years before the test. One parent argued, "I think our son has become too irresponsible since studying for INUEE. He's always having another person in the family do his jobs. He even doesn't do some basic things like getting a glass of water to drink." Another parent mentioned, "my son has never been such impatient and over-sensitive. He's always making unreasonable demands on us and he thinks because he's preparing for INUEE, we must move heaven and earth for him. We have always thought that parenting means providing all the conditions for our children to study, but unfortunately, despite providing the conditions, INUEE has negatively affected his upbringing."

Another question parents were asked concerned whether they were under any kind of pressure from family members, different people, or even institutes regarding their children's performance in INUEE. Different sources of pressure were named, and some of the parents named more than a single pressure source. From a total of 27 parents interviewed, 18 (66.66%) parents stated that they were under the pressure of their spouse, 17 (62.96%) parents also expressed that they were under the pressure of their relatives, and 6 (22.22%) parents stated that they were under the pressure of others (e.g., colleagues, neighbors) as well. The majority of parents referred to the pressure put on them by their spouses for providing a stress-free environment for their children while preparing for INUEE. One parent said, "my husband is always telling me to provide the best condition for our daughter to study for INUEE. Don't make her do any extra activity like doing housework, going shopping etc." Most parents also mentioned that most of their acquaintances have set high expectations for their children's performance in INUEE and the fear of failing to meet such expectations puts the family under too much psychological pressure. One parent expressed, "my relatives are always asking us about our

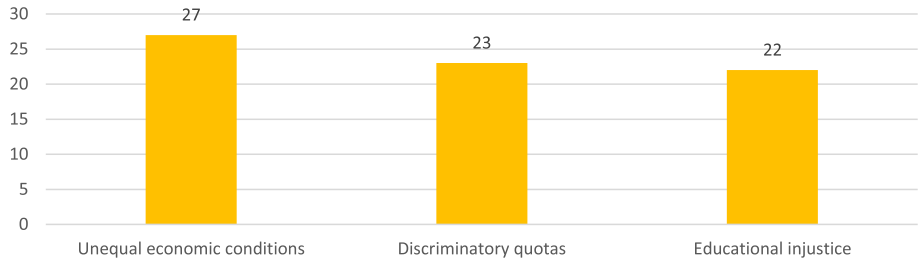

Fig. 6 Social discrimination types as a consequence of INUEE

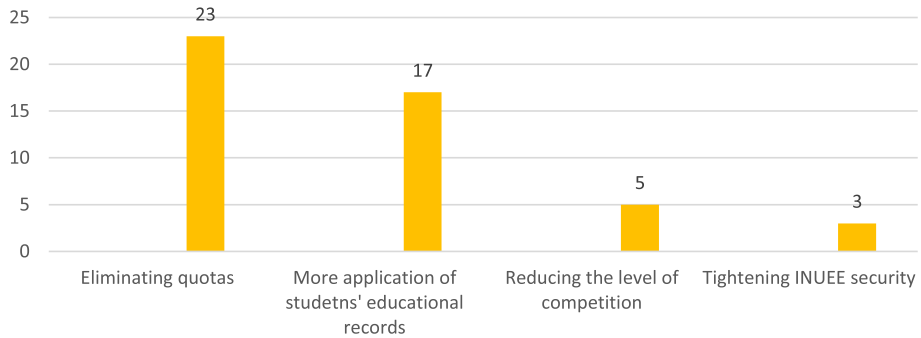

Fig. 7 Parents' suggestions for change in INUEE

daughter's level of preparation for INUEE. They're always asking about why her studying for INUEE won't finish. They also call her by some titles like 'Doctor,' maybe sarcastically. Such over-attention on the part of our relatives have put our family under too much pressure. We're always thinking what will happen if she fails the exam." They also referred to other people's sarcastic remarks such as "aren't you done with INUEE yet?" which were reportedly too stressful and even disappointing for them and their children. In addition, some parents expressed that comparing their children's performance with others by their acquaintances puts a severe psychological pressure on the family.

Furthermore, the parents were asked if any social discrimination occurs as a consequence of INUEE results. As demonstrated in Fig. 6, all 27 (100%) parents referred to economic discrimination, indicating that students in lower- and middle-class families were being discriminated against as a consequence of not having sufficient financial resources to enroll in INUEE preparation classes and buying INUEE preparation textbooks. Also, 23 (85.18%) parents stated that INUEE quotas were discriminatory and unfair. Furthermore, 22 (81.48%) parents referred to educational injustice created as a consequence of unequal environmental and educational facilities.

Finally, the parents were asked if they had any suggestion(s) for change in INUEE. The results obtained from the analysis of the responses indicated that the majority of the parents believed that a single session 4-h test for measuring a 12-year-long learning process was unfair. The results of the analysis of the parents' suggestions are summarized in Fig. 7.

As Fig. 7 indicates, 23 (85.18%) parents suggested eliminating the INUEE quotas because they deeply believed that applying quotas to students' INUEE scores was unfair. In addition, 17 (62.96%) suggested that the test takers' educational records need to be more influentially considered as a part of their INUEE scores, 5 (18.51%) suggested reducing the

competitiveness level of the test, and 3 (11.11%) had concerns about INUEE security and called for enhanced security measures for safeguarding INUEE questions leakage before the test.

Discussion

Consequential validity (Messick, 1989) of a test used for certain purposes entails delving into the perceptions of the stakeholders of a high-stakes test like INUEE. In the Iranian system of general education, INUEE is a high-stake gatekeeping test that is annually administered at a nationwide scale to over a million Iranian high school graduates for the selection of candidates for more than hundreds of higher education academic programs. It is quite evident that such competitive high-stakes test will have grave impacts on the stakeholders' lives. Moreover, any attempt to ameliorate the fairness of the decisions passed based on such tests' results needs an awareness of the perceptions of the most important stakeholders of the test including the test takers themselves, the test developers, teachers, and even test takers' family members. On this basis, the purpose of this study was to privilege students' and parents' voices in order to elicit their thoughts and feelings with respect to INUEE.

Analysis of the participants' opinions reflected in both the INUEE Impact Questionnaire and the structured interview indicated that the majority of the participants were negatively affected by INUEE. From students' vantage points, INUEE had negative impacts on their physical and psychological well-being, familial and social relationships, educational goals, economic status, entertainments, motivation, and self-confidence. Furthermore, analysis of parents' interview transcripts indicated that not only students but their parents also fall prey to INUEE consequences. Family relationships and entertainments, families' economic status, and students' upbringing are all negatively affected by INUEE.

The attested finding that high-stakes testing has negative impacts on students' well-being aligns with some previous studies such as Larson et al. (2010), McCaleb-Kahan and Wenner (2009), and von der Embse (2008), who found that after testing, some overly anxious children displayed symptoms like crying, illness, and outbursts of anger. This finding is also in line with Triplett and Barksdale (2005) and Triplett et al. (2003), in which students reported negative impacts such as fear, anxiety, stress, physical illness, and powerlessness.

In addition, the negative social impact of high-stakes testing verified in this study lends support to the results of Valenzuela (2000), who voices concerns about the objectification of students and the negation of students as individuals and as cultural beings with distinct experiences, needs, and desires that accompany their differences.

Regarding the negative educational impact of INUEE, our findings are consistent with some previous studies such as Barksdale and Thomas (2000), Connor (2002), Westfall (2010), and Wyn et al. (2014) who also reported that the parents had negative attitudes about the high-stakes tests, saw little value in them, did not perceive the tests as an appropriate and fair measure of their children's achievements and skills, and believed that a single test could not measure how satisfactorily students perform.

Studies have also confirmed the current study findings and concerns about the fairness of high-stakes tests as a measure of students' achievements and abilities. For instance, Moon et al. (2007), Rapp (2002), Reichel (2009), Smith (1991), and Wright and Choi (2006) confirm the negative educational impacts of high-stakes tests.

Furthermore, the verified negative impact of INUEE on students' motivation is consistent with the findings of studies such as Clemmitt (2007), Jones et al. (2003), Mesler (2008), Nichols and Berliner (2008), and Valli et al. (2008). Swain et al. (2018) also showed that students' choice for adopting a curriculum, assessment, and pedagogy focused on NAPLAN, a high-stakes test in Australia, gave rise to students experiencing negative reactions such as fear, sadness, anxiety, and panic.

Our findings concerning the negative impacts of high-stakes testing on familial relationships, family recreational activities, and the economic status of the family are also significant in their own right as they suggest how the unintended consequences of high-stakes testing can extend beyond schools and into not only the homes of students but also into the families' social lives.

The findings suggest that in many cases, INUEE contradicts the basic elements of educational justice. For instance, from an economic perspective in particular, the findings indicated that some parents had taken out loans, borrowed money, and even sold their belongings to be able to afford the expenses associated with INUEE textbooks, Supplementary materials, and the costs for private schools and institutes offering high-quality teach-to-the-test trainings. This implies that students with lower socioeconomic backgrounds were to be put at a great disadvantage in such an unfair competitive situation.

Another instance of the lack of compliance with the basic principles of quality education for all was found to be the allocation of quotas to some special groups of students. This reflects the existing bias and unfairness in such a situation on the grounds that an individual's group membership determines that s/he achieves significantly differently from the whole population. In addition, confirming some previous studies' results (e.g., Macqueen et al., 2018) as well as the current study findings concerning the breach of educational justice by the high-stakes tests like INUEE, the Iranian Constitution (article 19, 30) also postulates that "the people of Iran enjoy equal rights, regardless of the tribe or ethnic group to which they belong. Color, race, language, and other such considerations shall not be grounds for special privileges." Also, the discriminatory quotas, the unequal economic conditions, and the unequal educational facilities that parents referred to in this study testify to the ethical concerns raised by Shohamy (1993; 1997) which indicated that tests have ethically questionable and unstated political purposes that might be completely different from their stated purposes and Lynch (1997) who referred to much broader ethical considerations of test use and questioned the bases on which we can or cannot defend the uses we make of the tests. These authors have also called for testers and individuals who are impacted by test use, to critically analyze the purposes for which tests are utilized and to ensure that the test uses are fair and ethical. Thus, it is apparent that there is an urgent need for policymakers to take into consideration how their policies regarding high-stakes testing in general and INUEE in particular put potential test takers and

their families at a great disadvantage, particularly those families who are at a lower socioeconomic status, and to adjust educational policies accordingly.

These findings also cast doubt on the usefulness of INUEE, as a nationwide university selection test, and suggest that it is undermining the quality of education received by Iranian high school students. In order to appropriately respond to the call for accountability and professionalism, it behooves policymakers, educators, principals, and all other decision-makers to first raise their awareness with respect to such unintended consequences and then acknowledge and reconsider their educational policies in order to minimize the negative consequences of INUEE on test takers and their families. As Kunnan (2004) puts it, those responsible for the design and use of a test, i.e., developers, providers, and policymakers, need to make sure that the instrument does no harm. Kunnan's idea in this regard reminds Spolsky's warning that "tests should be labeled just like dangerous drugs: 'Use with care!'" (Spolsky, 1981, p. 20).

Conclusion and implications

The present study aimed to scrutinize the impacts of INUEE on Iranian high school students and their parents. The findings revealed that the students and their parents were negatively affected by INUEE, which indicates that the current implementation of INUEE as a high-stakes university entrance selection test is in conflict with the basic principles of educational justice and even its own intended selection purpose, making it an impractical instrument for screening the most capable high school graduates and placing them in the most relevant and desirable academic programs. Thus, even if INUEE designers might have intended to boost the quality of education in Iran, their efforts, as revealed by the findings of the present study, have apparently backfired. Technically speaking, the findings of the study put the consequential validity, usefulness, and fairness of INUEE under serious question.

On the basis of the obtained results, policymakers are strongly advised to either consider eliminating the current testing system or changing their educational policies so as to minimize the negative consequences of INUEE on potential test takers and their families in an attempt to make it fairer. In practical terms, they are called for a policy shift that minimizes the threats of INUEE, enhances opportunities for improving the quality of education, and addresses the need to enhance the test takers' overall well-being.

In addition, school principals need to seek ways for engaging not only students from different educational levels and backgrounds, but also their parents as active participants in the education process. Through parents' engagement in the decision-making process and using their children's knowledge in educational decisions (Pushor, 2007), principals can gain a rich opportunity to delve into students' educational lives in schools and explore what goes on in students' minds behind classroom doors (Swain et al., 2018). High school teachers also need to recognize the consequences of high-stakes tests and offer students strategies for handling them. Although this might add to their workload, teachers are in a position to instill self-confidence and motivation in students.

In closing, the findings substantiated the need for a paradigm shift by which the current one-size-fits-all INUEE is substituted with a more humane, fair, and unbiased assessment system that minimizes the negative consequences, unfairness, and the involved reported bias. However, the study findings need to be interpreted cautiously as

a number of limitations might restrict the generalizability of the study results. One of the limitations of the study was that the data obtained from students contained only quantitative data. It would add more flesh to the bone if we were able to elicit some qualitative data from a subset of our student sample. Another limitation of the study was that it focused only on three western provinces of Iran, especially with regard to the fact that the samples were selected through a convenience sampling procedure, which might not produce fully representative results.

Finally, it needs to be acknowledged that a more vivid picture of the impacts of INUEE would have been resulted if other stakeholders' viewpoints were also considered; however, space limitation of the present manuscript excluded the possibility of the inclusion of high school teachers' and principals' perspectives. The perspectives of these two groups of stakeholders are presented elsewhere (Ahmadi Safa & Sheykholmoluki, 2023), and interested readers are referred to it.

Abbreviations

INUEE	Iranian National University Entrance Examination
TOA	Theory of action
PCA	Principal component analysis
KMO	Kaiser-Meyer-Olkin
NAPLAN	National Assessment Program-Literacy and Numeracy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40468-023-00254-0.

Additional file 1:Appendix A. Iranian National University Entrance Exam (INUEE) Impact Questionnaire. **Appendix B.** Interview questions for parents. **Appendix C.** The results of rotated factor matrix (PCA with Varimax rotation) and communalities. **Appendix D.** Descriptive statistics for the items of the INUEE questionnaire.

Acknowledgements
Not applicable.

Authors' contributions
Both authors equally contributed to the completion of the study. A substantial contribution was made by MAS in the data collection, data analysis, and data interpretation as well as the composition of the final manuscript. HS also made his equal contribution to the study during different phases of the investigation. Both authors also read and approved the final manuscript.

Funding
The project was funded by Bu-Ali Sina University, Iran.

Availability of data materials
The authors state that all the data supporting the findings of this study are available within the article.

Declarations

Ethics approval and consent to participate
Ethics approval is not applicable in the context of the study.
Informed consent of all participants was obtained at the outset of the study.

Consent for publication
Both authors express their consent for the publication of the present article.

Competing interests
The authors declare that they have no competing interests.

Received: 17 June 2023 Accepted: 7 August 2023
Published online: 25 August 2023

References

Ahmadi Safa, M., & Sheykholmoluki, H. (2023). The impacts of a nationwide high-stakes test from high school teachers and principals' perspectives: a qualitative study. *International Journal of Language Testing,13*(1), 104–132. https://doi.org/10.22034/ijlt.2022.352176.1179

Ajideh, P., & Mahmoudi, L. (2017). Iranian pre-university students' perceptions towards the English section of the Iranian National University Entrance Exam (INUEE): a washback study. *Journal of English Language Teaching and Learning,20*, 25–48.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics,14*, 115–129. https://doi.org/10.1093/applin/14.2.115

Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership,60*(5), 32–38.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly,2*, 1–34. https://doi.org/10.1207/s15434311laq0201_1

Barksdale, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: teachers and parents speak out. *Journal of Teacher Education,51*(5), 384. https://doi.org/10.1177/0022487100051005006

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives.,8*, 70–91. https://doi.org/10.1080/15366367.2010.508686

Bennett, R. E., Kane, M. T., & Bridgeman, B. (2011). *Theory of action and validity argument in the context of through-course summative assessment*. Paper presented at invitational Research Symposium on Through Course Summative Assessment, Atlanta, GA. Retrieved August, 18, 2013 from www.k12center.org/rsc/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf.

Booher-Jennings, J. (2008). Learning to label: Socialization, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education,29*(2), 149–160. https://doi.org/10.1080/01425690701837513

Bracey, G. (2003). *On the death of childhood and the destruction of public schools*. Heinemann.

Brady, A. L. (2008). Effects of standardized testing on teachers' emotions, pedagogy and professional interactions with others. *ETD Archive* (p. 39)

Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 1–16). Praeger.

Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics,29*, 118–131. https://doi.org/10.1017/S0267190509090102

Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Horberger (Eds.), *Encyclopedia of language and education vol. 7: Language testing and assessment* (2nd ed., pp. 349–64). Dordrecht: Springer. https://doi.org/10.1007/978-0-387-30424-3_186

Clemmitt, M. (2007). Students under stress. *CQ (Congressional Quarterly) Researcher,17*(25), 577–600.

Connor, F. P. (2002). *Parents' perceptions regarding the Virginia standards of learning* (Unpublished doctoral dissertation), University of Virginia, Charlottesville.

Davies, A. (2008). Ethics, professionalism, rights and codes. In E. Shohamy & N. H. Hornberger (Eds.), *Language testing and assessment. (Volume 7 of Encyclopedia of Language and Education* (2nd ed., pp. 429–443). Dordrecht: Springer.

Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Journal of Studies in Educational Evaluation,64*(33), 1–11.https://doi.org/10.1016/j.stueduc.2019.100824.

Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Fitzgerald, L. (2015). Consequences of high-stake testing. *Education Masters* (p. 306)

Fulcher, G. (2014). Philosophy and language testing. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 1431–1451). John Wiley & Sons.

Ghorbani, M. R. (2008). *Washback effect of the university entrance examination on Iranian pre-university English language teachers' curriculum planning and instruction (Doctoral dissertation)*. Malaysia: University of Putra.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.https://doi.org/10.4324/9780203793206.

Guloglu-Demir, C., & Kaplan-Keleş, O. (2021). The impact of high-stakes testing on the teaching and learning processes of mathematics. *Journal of Pedagogical Research,5*(2), 119–137.https://doi.org/10.33902/JPR.2021269677.

Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Re Anderson multivariate data analysis: A global perspective*. Pearson Prentice Hall.

Hazel, C. (2010). Interactions between bullying and high-stakes testing at the elementary school level. *Journal of School Violence,9*(4), 339–356. https://psycnet.apa.org/doi/10.1080/15388220.2010.507142 .

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research,103*, 219–230. https://psycnet.apa.org/doi/10.1007/s11205-011-9843-4 .

Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Rowman and Littlefield Publishers Inc.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement,50*, 1–73.https://doi.org/10.1111/jedm.12000.

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport: American Council on Education and Praeger.

Kunnan, A. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge University Press.

Landry, D. (2006). Teachers' (K-5) perceptions of student behaviors during standardized testing. *Curriculum & Teaching Dialogue,8*(1/2), 29–40.

Larson, H. A., El Ramahi, M. K., Conn, S. R., Estes, L. A., & Ghibellini, A. B. (2010). Reducing test anxiety among third grade students through the implementation of relaxation techniques. *Journal of School Counseling,8*(19), 1–19.

Lynch, B. (1997). In search of the ethical test. *Language Testing,14*, 315–27.https://doi.org/10.1177/026553229701400308.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing,18*, 351–372.
Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates Publishers.
Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2018). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Journal of Language Testing,36*(2), 265–287.https://doi.org/10.1177/0265532218775758.
Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education* (pp. 83–121). University of Chicago Press.
Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high-stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Information Age Publishing Inc.
McCaleb-Kahan, P., & Wenner, R. (2009). *The relationship of student demographics to 10th grade MCAS test anxiety.* Paper presented at the 40th Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT, October 2009. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1002&context=nera_2009.
McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly,3*, 31–51.https://doi.org/10.1207/s15434311laq0301_3.
McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.
Mesler, A. (2008). *High stakes and higher impacts: The effects of high-stakes testing on school and classroom operations*. Durham: Duke University. Retrieved from: http://uwp.duke.edu/uploads/assets/Mesler_09.pdf .
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education & Macmillan.
Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist,35*(11), 1012–1027. https://psycnet.apa.org/doi/10.1037/0003-066X.35.11.1012 .
Moon, T. R., Brighton, C. M., Jarvis, J. M., & Hall, C. J. (2007). *State standardized testing programs: Their effects on teachers and students. (RM07228)*. Storrs: The National Research Center for the Gifted and Talented.
Moses, M. S., & Nanna, M. J. (2007). The testing culture and the persistence of high stakes testing reforms. *Education and Culture,23*(1), 55–72.https://doi.org/10.1353/eac.2007.0010.
Nichols, S. L., & Berliner, D. C. (2008). Testing the joy out of learning. *Educational Leadership,65*(6), 14–18.
Pallant, J. F. (2000). Development and validation of a scale to measure perceived control of internal states. *Journal of Personality Assessment,75*(2), 308–337.https://doi.org/10.1207/S15327752JPA7502_10.
Pushor, D. (2007). *Parent engagement: Creating a shared world*. Paper Presented at the Ontario Education Research Symposium. Retrieved March 3, 2010 fromhttp://www.edu.gov.on.ca/eng/research/pushor.pdf.
Rapp, D. (2002). National Board Certified teachers in Ohio give state education policy, classroom climate, and high-stakes testing a grade of F. *Phi Delta Kappan,84*(3), 215–218. https://doi.org/10.1177/003172170208400309.
Reichel, A. G. (2009). Unintended consequences: The teacher's story. In T. A. Price & E. Peterson (Eds.), *The myth and reality of no child left behind: public education and assessment.* Lanham: University Press of America, Inc.
Rezaeian, M., Seyyedrezaei, S. H., Barani, G., & Seyyedrezaei, Z. S. (2020). Construction and validation of educational, social and psychological consequences questionnaires of EPT as a high-stakes test. *International Journal of Language Testing,10*(2), 33–70.
Salehi, H., & Yunus, M. (2012). The washback effect of the Iranian universities entrance exam: teachers' insights. *Journal of Language Studies,12*(2), 608–628.
Shepard, L. (2003). The hazards of high-stakes testing. *Issues in Science and Technology,19*(2), 53–58.
Shohamy, E. (1993). *The power of test: The impact of language testing on teaching and learning*. National Foreign Language Center Occasional Papers.
Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? *Language Testing,14*, 340–49.
Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson.
Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher,20*(5), 8–11.https://doi.org/10.3102/0013189X020005008.
Spolsky, B. (1981). Some ethical questions about language testing. In C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and problems in language testing* (pp. 5–21). Peter Lang.
Stecher, B. M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment,74*, 259–281.
Swain, K., & Pendergast., D., & Cumming., J. (2018). Student experiences of NAPLAN: Sharing insights from two school sites. *The Australian Educational Researcher,45*(3), 315–342.https://doi.org/10.1007/s13384-017-0256-5.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson Education. Inc.
Thomas, R. M. (2005). *High-stakes testing: Coping with collateral damage.* Lawrence Erlbaum Associates Publishers.https://doi.org/10.4324/9781410612809.
Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research,37*(2), 237–260.https://doi.org/10.1207/s15548430jlr3702_5.
Triplett, C., Barksdale, M. A., & Leftwich, P. (2003). Children's perceptions of high stakes testing. *Journal of Research in Education,13*(1), 15–21.
Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing,39*(2), 212–238.https://doi.org/10.1177/02655322211050600.
Valenzuela, A. (2000). The significance of the TAAS test for Mexican immigrant and Mexican American adolescents: a case study. *Hispanic Journal of Behavioral Sciences,22*(4), 524.https://doi.org/10.1177/0739986300224009.
Valli, L., Croninger, R. G., Chambliss, M. J., Graeber, A. O., & Buese, D. (2008). *Test driven: high-stakes accountability in elementary schools*. Teachers College Press.
von der Embse, N. P. (2008). *High stakes testing, socio-economic status and test anxiety: Opportunities for impact*. Oxford: Thesis submitted to the Department of Educational Psychology, Miami University. Retrieved from http://etd.ohiolink.edu/send-pdf.cgi/vonderEmbseNathaniel.pdf?miami1214510534.
Walker, T. (2014). NEA survey: Nearly half of teachers consider leaving profession due to standardized testing. *National education association* (p. 16)

Watson, C., Johanson, M., Loder, M., & Dankiw, J. (2014). Effects of high-stakes testing on third through fifth grade students: Student voices and concerns for educational leaders. *Journal of Organizational Learning and Leadership,12*(1), 1–11.
Westfall, D. M. (2010). *Parental perceptions of the effects of the high-stakes TAKS test on the home lives of at-risk fifth grade students* (unpublished doctoral dissertation), University of Houston, Houston.
Wright, W. E., & Choi, D. (2006). The impact of language and high-stakes testing policies on elementary school English language learners in Arizona. *Education Policy Analysis Archives,14*(13), 1–71. https://doi.org/10.14507/epaa.v14n13.2006.
Wyn, J., Turnbull, M., & Grimshaw, L. (2014). *The experience of education: The impacts of high stakes testing on school students and their families. A qualitative study*. Whitlam Institute.
Ysseldyke, J., Nelson, J., Christenson, S., Johnson, D., Dennis, A., Triezenberg, H., & Hawes, M. (2004). What we know and need to know about the consequences of high-stakes testing for students with disabilities. *Council for Exceptional Children,71*(1), 75–94.https://doi.org/10.1177/001440290407100105.

Publisher's Note