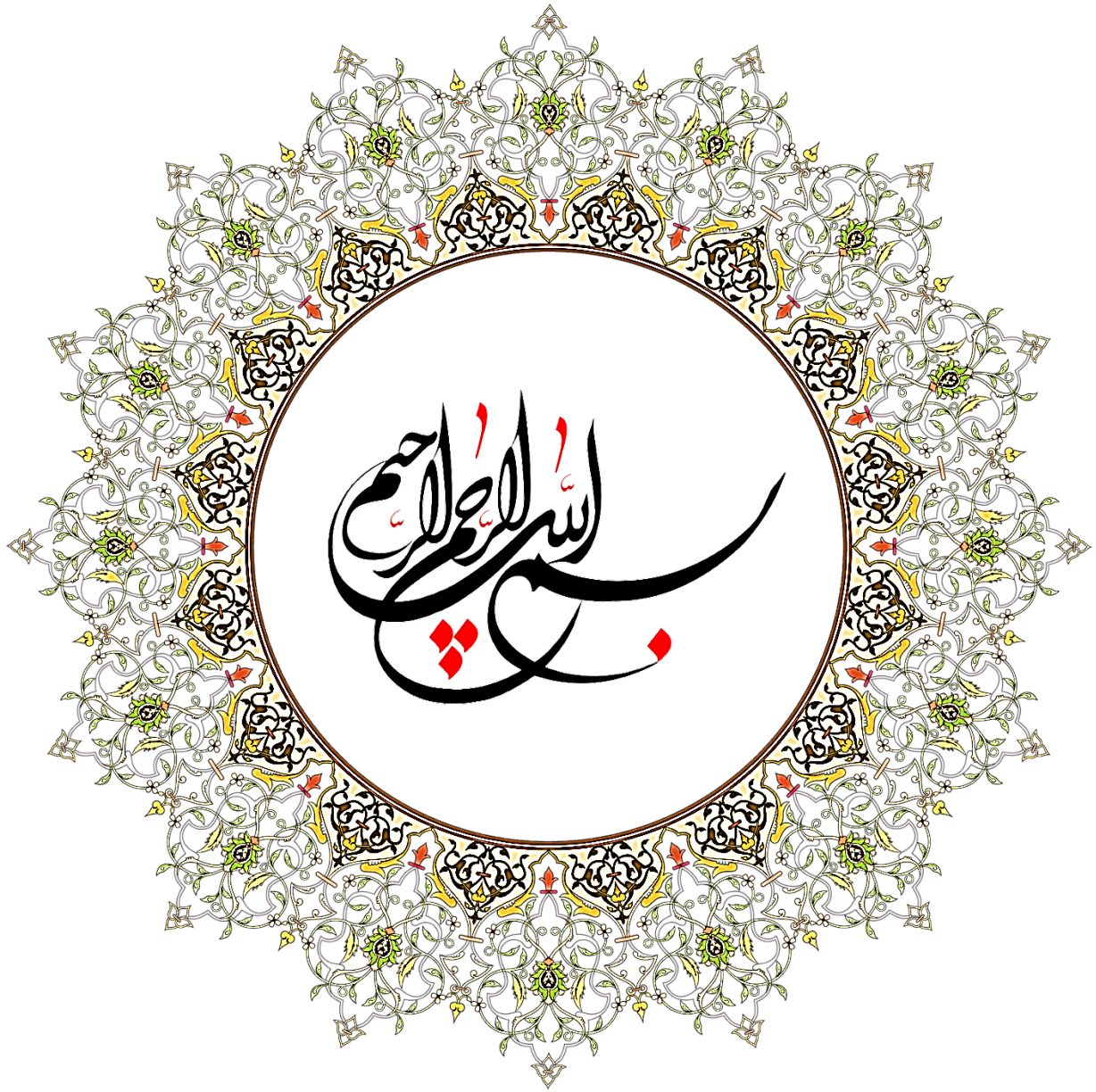
بسم الله الرحمن الرحیم

دانشگاه شهید چمران اهواز

تدوین و نگارش متون علمی

مجتبی جهانی فر
دانشکده روانشناسی و علوم تربیتی
دانشگاه شهید چمران اهواز
نیم سال اول سال تحصیلی ۹۹-۹۸

درس ششم

اجزای متون علمی
این جلسه (بحث)

بحث و نتیجه گیری Conclusion

- "بحث" یک قسمت فنی و تخصصی در تحقیق و نگارش علمی است.
- "بحث" یعنی معنای یافته های تحقیق از زبان پژوهشگر.
- "بحث" بر مبنای استناد و استدلال بنا می شود.

- در بحث شما به جایی رسیدید که می توانید نتایج را ارزیابی، تفسیر و توصیف کنید.
- نتیجه تفسیر و ارزیابی به صورت استنباط و نتیجه گیری خودش را نشان می دهد.
- بحث باید با بیان واضحی از فرضیه ها اصلی تایید شده و تایید نشده آغاز کنید.
- اگر فرضیه ای تایید نشده دارید، باید توضیحات تعقیبی را پیش بگیرید.
- تفاوت ها و شباهت های بین نتایج شما و دیگران باید به منظور روشن شدن و تایید نتایج خودتان مورد استفاده قرار گیرد.
- تفسیر شما از نتایج باید مواردی همچون : منابع سوگیری، عدم دقت اندازه ها، تعداد و هم پوشی آزمون ها، اندازه های اثر، محدودیت ها و ضعفهای پژوهش.
- محدودیت های پژوهش خود را ذکر کنید و به تبیین جایگزین ها بپردازید.
- بحث را با تفسیر مستدل و موجه اهمیت یافته ها به پایان برسانید.

برای نوشتن بحث طرح واره زیر را در ذهن خود مرور کنید

مرور یافته اصلی تحقیق بدون ذکر عدد یا نتیجه آزمون آماری

خلاصه نتایج محققان دیگر که همسو با نتایج این تحقیق هستند

خلاصه نتایج محققان دیگر که مغایر نتایج این تحقیق هستند
(استدلال محقق در مورد دلایل این مغایرتها).

اهمیت یافته ها و اینکه چرا خوانندگان باید به این نتایج توجه کنند

محدودیت های تحقیق و اثر آنها بر نتایج تحقیق و تعمیم پذیری یافته ها

نتیجه گیری نهایی و پیامدهای نظری و عملی پژوهش

کاربرد های نتایج تحقیق و اینکه آیا این کاربرد ها تضمین شده است؟

کدام مشکلات و مسائل حل نشده باقی مانده اند و چه سوالهایی بر مبنای این پژوهش مطرح می شود؟

پیشنهاد برای پژوهش های بعدی با توجه به یافته ها.

در نهایت در بخش بحث باید با توجه به نتایجی که ارائه کرده است پاسخی قاطع به پرسش های پژوهش داده شود و مشخص گردد که تحقیق انجام شده چیزی را به مجموعه علم اضافه کرده است.

Journal of Educational Measurement
Spring 2019, Vol. 56, No. 1, pp. 28–50

A Method for Detecting Regression of Hard and Easy Item Angoff Ratings

Adam E. Wyse
Renaissance
Ben Babcock
The American Registry of Radiologic Technologists

One common phenomenon in Angoff standard setting is that panelists regress their ratings in toward the middle of the probability scale. This study describes two indices based on taking ratios of standard deviations that can be utilized with a scatterplot of item ratings versus expected probabilities of success to identify whether ratings are regressed in toward the middle of the probability scale. Results from a simulation study show that the standard deviation ratio indices can successfully detect ratings for hard and easy items that are regressed in toward the middle of the probability scale in Angoff standard-setting data, where previously proposed indices often do not work as well to detect these effects. Results from a real data set show that, while virtually all raters improve from Round 1 to Round 2 as measured by previously developed indices, the standard deviation ratios in conjunction with a scatterplot of item ratings versus expected probabilities of success can identify individuals who may still be regressing their ratings in toward the middle of the probability scale even after receiving feedback. The authors suggest using the scatterplot along with the standard deviation ratio indices and other statistics for measuring the quality of Angoff standard-setting data.

An important step in developing many examinations is determining cut scores to classify examinees. In credentialing, one typically determines cut scores to separate examinees who possess the required knowledge, skills, and abilities for a credential from those who do not. In an educational testing, one often identifies several cut scores to separate examinees into multiple performance categories. One common approach to determine such cut scores is the Angoff (1971) standard-setting method. The Angoff method uses a group of panelists to estimate the probability that minimally competent examinees would answer each item correctly. These probability ratings are then analyzed to determine a cut score for each panelist and the mean or median of the panelist cut scores is taken as the estimated group cut score.

There are a few aspects of the Angoff method that warrant additional discussion. First, a panelist's task is to consider the content of each item, the nature of practice, and the definition of the minimally competent examinee and give probability ratings that are consistent with how they think minimally competent examinees would perform. Panelists may differ on how they think minimally competent examinees would perform and give different ratings, which implies that there is not an expectation that all panelists give ratings that are indicative of the same cut score. However, there is typically an expectation that the panelists' ratings should align with the level of performance that they think a minimally competent examinee would demonstrate.

© 2019 by the National Council on Measurement in Education

مشخصات مجله

عنوان مقاله

نویسندگان

چکیده

شروع متن با مقدمه

Quarterly of Educatinal Psychology
Allameh Tabataba'i University
Vol. 14, No.50, Winter 2019

فصلنامه روان‌شناسی تربیتی
دانشگاه علامه طباطبائی
سال چهاردهم، شماره پنجاه، زمستان ۹۷، ص ۲۵ تا ۵۴

اثربخشی آموزش ذهن‌آگاهی بر اضطراب امتحان و نگرش نسبت به مدرسه در دانش آموزان پایهٔ یازدهم

فاطمه خرمی[۱]، علی اکبر سیف[۲]، علیرضا کیامنش[۳]، فریبرز درتاج[۴]

تاریخ وصول: ۱۳۹۷/۰۱/۱۰ تاریخ پذیرش: ۱۳۹۷/۰۵/۲۹

چکیده

هدف پژوهش حاضر بررسی اثربخشی آموزش ذهن‌آگاهی در کاهش اضطراب امتحان و افزایش نگرش نسبت به مدرسه در دانش‌آموزان پایهٔ یازدهم شهر تهران و روش پژوهش از نوع آزمایشی با طرح پیش‌آزمون-پس‌آزمون با گروه کنترل بود. جامعهٔ آماری پژوهش شامل کلیهٔ دانش‌آموزان پایهٔ یازدهم شهر تهران، در سال تحصیلی ۹۷-۱۳۹۶، بودند. از جامعهٔ مذکور، ۸۸ دانش آموز (۴۴ دختر و ۴۴ پسر) با استفاده از روش خوشه‌ای چندمرحله‌ای که نمراتِ بالایی در پرسشنامهٔ اضطراب امتحان فریدین (۱۹۹۷) و نمراتِ پایینی در پرسشنامهٔ نگرش به مدرسهٔ مک‌کوچ و سیگل (۲۰۰۳) کسب کرده بودند، انتخاب و به تصادف در ۴ گروه (۲ گروهِ آزمایش و ۲ گروه کنترل) جایگزین شدند. سپس گروه‌های آزمایشی، در ۸ جلسهٔ ۲ ساعته، پروتکل ذهن‌آگاهی را دریافت کردند. پس از اتمام جلسات، تمامی آزمودنی‌ها مجدداً مورد ارزیابی قرار گرفتند. داده‌ها با استفاده از روشِ آماری تحلیل کوواریانس تک متغیری و چند متغیری تجزیه‌وتحلیل شدند و یافته‌های پژوهش نشان داد که آموزش ذهن‌آگاهی در کاهشِ اضطراب امتحان و افزایشِ نگرش به مدرسه تأثیرگذار است.

واژگان کلیدی: اضطراب امتحان، ذهن‌آگاهی، نگرش نسبت به مدرسه، دانش‌آموزان.

۱. دانشجوی دکتری روان‌شناسی تربیتی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران (نویسنده مسئول) anooshe.khorami@gmail.com
۲. استاد، گروه روان‌شناسی تربیتی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران
۳. استاد، گروه روان‌شناسی تربیتی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران
۴. استاد، گروه روان‌شناسی تربیتی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

The second part of the Angoff method is translating panelist ratings into cut scores. When applying item response theory (IRT) models, one uses an equating method to place the ratings onto the same scale as the examinee item response data and to estimate cut scores using an IRT estimator (see Wyse, 2017). The equating process is usually conducted by fixing the item parameters to the values estimated from item response data. Fixing the item parameters assumes that panelist ratings should align with the expected probabilities of success on the items based on the rater's estimated cut score and the IRT model (e.g., item characteristic curve values at the estimated cut scores) or there may be inaccuracies in the estimated cut score. The idea that Angoff ratings should align with expected probabilities based on the rater's estimated cut score and the IRT model is the foundation for Reckase charts (Reckase, 2001), which are sometimes provided as rater feedback, as well as indices often used to evaluate ratings (see Hurtz & Jones, 2009). These observations provide a theoretical foundation for anticipating that Angoff ratings should align with expected probabilities of success based on estimated cut scores and the IRT model used on the exam.

Research, however, suggests that many challenges can occur when performing the Angoff method and that ratings may not always align with what one may expect (see Brandon, 2004; B. E. Clauser, Mee, Baldwin, Margolis & Dillon, 2009; B. E. Clauser, Mee, & Margolis, 2013; B. E. Clauser, Swanson, & Harik, 2002; Humphry, Heldsinger, & Andrich, 2014; Hurtz & Auerbach, 2003; Hurtz & Jones, 2009; Margolis & Clauser, 2014; Reckase, 2001; Shepard, 1995; Shephard, Glader, Linn, & Bohrnsted, 1993; Taube, 1997; Wyse, 2013, 2018; Wyse & Reckase, 2012). One prevalent challenge is that panelists often regress item ratings in toward the middle of the probability scale, thinking that hard items are easier and easy items are more difficult than they are compared to expected probabilities of success based on their estimated cut scores and the IRT model. Shephard and colleagues (Shepard, 1995; Shepard et al., 1993) first reported this in National Assessment Educational Progress (NAEP) standard settings in the early 1990s. Schulz (2006) later reproduced this finding in an NAEP mathematics standard setting. The pattern of regressing Angoff ratings in toward the middle of the probability scale for hard and easy items is not isolated to NAEP. Taube (1997) analyzed data from four credentialing programs, finding that panelists thought hard items were too easy and easy items were too hard in each program. Similarly, Wyse (2018) observed this phenomenon in two credentialing program standard settings. Humphry et al. (2014) also found these effects in a K–12 standard setting in Australia.

Figure 1 shows this phenomenon for a rater from an actual standard setting. The diagonal identity line shows the ideal scenario where the ratings match the expected probabilities of success based on a rater's cut score and the IRT model. One would hope to find ratings that are close to the identity line, but that are randomly distributed above and below the identity line over the range of expected probabilities. The ratings in Figure 1 exhibit a linear trend with the hard items rated too easy and easy items rated too hard. It is important to note the systematic shifting of the points from the identity line, such that the Angoff ratings for harder items, which are toward the left of the figure, tend to be too high (i.e., above the identity line), and the Angoff ratings for easier items, which are toward the right of the figure, tend to be too low (i.e., below the identity line). It is also important to realize that this shift varies between

عنوان مکرر

شروع متن با مقدمه

پاورقی

مقدمه

اضطراب[1] به احساس ترس، نگرانی و ناراحتی که به‌واسطه تهدیدهای بالقوه بیرونی و درونی ایجاد شده است، اشاره دارد (گروپ و نیسچه[2]، ۲۰۱۳) و در نتیجه تنشی است که مردم مجبور هستند با آن در هر جنبه از زندگی و در هر سنی مبارزه کنند (لوفی، اوکاشا و کوهن[3]، ۲۰۰۴؛ تریفونی و شاهینی[4]، ۲۰۱۱). کودکان و نوجوانان در جریان رشد انواع گوناگون و طیف وسیعی از اضطراب‌ها را تجربه می‌کنند و گاه از چنان شدتی برخوردارند که زندگی روزمره و تحصیلی آنان را دشوار می‌سازد. یک طیف از این اضطراب‌ها که در ارتباط با عدم تمایل رفتن به مدرسه به کار برده می‌شود، اضطراب مدرسه[5] است. دانش‌آموز دارای اضطراب مدرسه ممکن است اضطراب اجتماعی، تحصیلی و یا هر دو را تجربه کند (پنا و لوساداً[6]، ۲۰۱۷).

یکی از انواع اضطراب مدرسه که به‌عنوان یک پدیدهٔ مهم و متداول آموزشی، رابطهٔ تنگاتنگی با عملکرد و پیشرفت تحصیلی دانش‌آموزان دارد، اضطراب امتحان[7] است. اضطراب امتحان یک رفتار درونی بین دانش‌آموزان و یک مشکل عاطفی مهم است که تأثیر منفی بر روی یادگیری داشته و به‌عنوان یک واکنش هیجانی یا حالتی از استرس در نظر گرفته می‌شود که قبل از امتحان و در زمان امتحان رخ داده و بسیاری از دانش‌آموزان آن را تهدیدی برای خود می‌دانند (حشمت[8] و همکاران، ۲۰۰۸). در خصوص پیشایندهای اضطراب امتحان پژوهشگران به عوامل مختلفی چون کمال‌گرایی (مازاکو[9]، ۲۰۰۷)، خودکارآمدی و خودپنداره ضعیف (کراس، کراچوک و راجانی[10]، ۲۰۰۸)، هدف‌گزینی (جی هانگ[11]، ۲۰۰۶)، جنسیت و پایهٔ تحصیلی (آیدین[12]، ۲۰۱۷) اشاره کرده‌اند (به نقل از

1. anxiety
2. Grupe & Nitschke
3. Lufi, Okasha & Cohen
4. Trifoni & Shahini
5. school anxiety
6. Pena & Losada
7. social anxiety
8. Hashmat
9. Mazzacco
10. Cross, Krachok & Rajani
11. Ghung
12. Aydin

عنوان مکرر

نمودار یا شکل

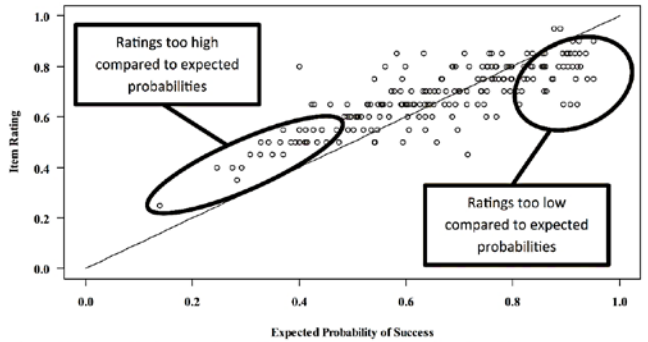

Figure 1. Example of regression of hard and easy item Angoff ratings for a rater.

raters. Some raters may give ratings that deviate more for hard items, others' ratings may deviate more for easy items, and others may have rating deviations for hard and easy items that are about the same. It is also possible for a rater to provide ratings that closely match the expected probabilities.

There are many reasons why ratings may deviate from expected values. In particular, training for the Angoff method often emphasizes the need to consider the appropriateness and relevance of items for minimally competent examinees, the content of the item, factors that may contribute to item difficulty, the difference between the overall difficulty of the item and the difficulty level of the item for minimally competent examinees, and how to understand and use feedback information. A panelist may believe that an item is inappropriate for minimally competent examinees and, thus, give a rating that is different from the expected probability of success. Alternatively, a panelist may think that an item tests extremely critical information that a minimally competent examinee must know and give a rating that does not align with the expected probability of success. Panelists may also focus on other factors and give unexpected item ratings. However, if a panelist's ratings are systematically off, as shown in Figure 1, it may indicate challenges in rating certain types of items and a different conceptualization of minimally competent examinee performance than that predicted by the IRT model used to score the exam.

The question then is how cut scores may be impacted if a panelist gives this pattern of ratings. Reckase (2006) explored the impact of regressing ratings in toward the middle of the probability scale and found potential cut score bias as a panelist's cut score moved away from an IRT θ value of 0. Wyse (2017) further explored these effects and found that regressing item ratings in toward the middle of the probability scale could bias cut scores. Using the Rasch model, Taube (1997) and Humphry et al. (2014) investigated whether rating hard items too high and easy items too low could change cut scores. Taube (1997) found that cut scores would be four to six questions lower and pass rates would be 7% to 15% higher if ratings were adjusted

تعبیر موجب خودانگیزشی[1] پایین و افزایش کناره‌گیری تحصیلی خواهد شد. آموزش‌های روان‌شناختی در راستای افزایش عزت‌نفس، کاهش اضطراب و توسعهٔ توانایی‌های شناختی روش مؤثری برای ایجاد نگرش مثبت نسبت به مدرسه محسوب می‌شود (استیل[2]، ۲۰۱۰). اضطراب امتحان با نگرش نسبت به مدرسه در ارتباط بوده و تعامل معناداری بین آن دو وجود دارد. هنگامی که اضطراب امتحان افزایش می‌یابد، دانش‌آموز بدون شک با نگرش منفی نسبت به مدرسه روبرو خواهد شد؛ بنابراین همبستگی معنادار بین نگرش به مدرسه و اضطراب امتحان باعث شده است تا پژوهشگران تلاش کنند تا اطلاعات دقیق‌تری را در مورد میزان همبودی دو سازه به دست آورند. شناسایی ارتباط بین اضطراب امتحان و نگرش به مدرسه می‌تواند اطلاعات فراوانی را اقدامات روانشناسی در محیط‌های آموزشگاهی ایجاد کند (بیابانگرد، ۱۳۸۷).

یکی از مداخلاتی که می‌تواند بر تعدیل هیجاناتِ حاصل از اضطراب امتحان و نگرشِ منفی به مدرسه تأثیر گذارد، آموزش ذهن‌آگاهی است (اکسلس و روزر[3]، ۲۰۰۹). تعاریف زیادی از ذهن آگاهی توسط پژوهشگران مختلف ارائه شده است؛ اما بر اساس نظر دان[4] (۲۰۱۰) همه این تعاریف دارای سه مؤلفهٔ مشترک می‌باشند. این سه مؤلفه عبارت‌اند از:

۱- ذهن آگاهی حالتی از هشیاری است که ممکن است بعضی افراد بیشتر از دیگران به آن دست یابند.

۲- افراد دارای این حالت هشیاری روی لحظه حال تمرکز می‌کنند.

۳- توجه افراد روی محرک‌های بیرونی (محیطی) و درونی متمرکز می‌شود.

ازاین‌رو، ذهن‌آگاهی به‌عنوان یکی از روش‌های درمانی است که می‌تواند هیجان‌های منفی از قبیل اضطراب، ترس و تنش را کاهش دهد و در رویکردهای رفتاردرمانی شناختی جای دارد (شیخ‌الاسلامی، ۱۳۹۲).

هدف مطلوب برنامهٔ کاهش استرس مبتنی بر ذهن آگاهی، پرورش آگاهی پایدار[5] و غیر واکنشی[6] در مورد تجربه‌های درونی (مثل شناختی، عاطفی و حسی) و تجربه‌های بیرونی

1. self-motivation
2. Steele
3. Eccles & Roeser
4. Dane
5. stedy
6. nonreactive

to Round 2 in the *b* metric but decreased in the *p* metric. These inconsistencies often occurred when panelists gave notably higher ratings in Round 2 compared to Round 1 with several extreme ratings. For example, Rater 3 had an SD_b of 1.08 in Round 1 and an SD_b of 1.27 in Round 2 compared to SD_ps of 1.10 and 1.08. This rater had zero ratings of 1.00 in Round 1 but eight such ratings in Round 2 with their mean rating increasing from .47 to .80. Similarly, Rater 10 had zero ratings of 1.00 in Round 1 and 50 ratings of 1.00 in Round 2 and their mean rating increased from .67 to .91. This rater had SD_bs of 1.13 and 1.19 and SD_ps of 1.00 and .86 in Rounds 1 and 2, respectively. The inconsistencies appeared to be a function of 1.00 ratings being set to missing when computing SD_b and the nonlinear relationship between the *b* and *p* metric being more pronounced with high ratings. In such cases, there is more variation in the unbounded *b* metric than in the bounded *p* metric. Second, one can see that there were only four instances where the standard deviation ratios were below 1.0, suggesting that it was common to have more variation in the expected probabilities than the item ratings.

Discussion

The systematic rating of hard items as too easy and easy items as too hard in comparison to values expected based on the rater's cut score and the measurement model applied on the exam is a complication in Angoff standard setting. This study showed that previously developed methods for evaluating Angoff ratings are not sensitive to panelists systematically regressing ratings for hard and easy items. In fact, the simulation study showed that many indices indicated that ratings were better or unchanged instead of signaling problems when ratings were regressed. Only the standard deviation ratio methods were consistently sensitive to giving higher ratings for hard items and giving lower ratings for easy items. The results of the real data study further demonstrated the utility that the scatterplot and standard deviation ratios could have in practice.

Standard setters should consider using the scatterplot and one of the new standard deviation indices in combination with other measures in this study to gain a more complete picture of the level of quality in Angoff ratings. For example, calculating the correlation between the ratings and the expected probabilities of success based on the panelists' estimated cut scores at the individual and group levels can indicate how well people are ordering the items. Pairing the correlation with a standard deviation ratio index and the scatterplot can give an idea of both the ordering and potential regression of hard and easy Angoff ratings. Pairing these indices can be particularly useful to detect regression of hard and easy items because the standard deviation ratios are measures of the slopes if one were to perform a linear equating to place the results onto the same scale and it is well known that equating methods tend to work better when correlations are high. Another application could be to pair the scatterplot and standard deviation ratios with the rater balance index as the rater balance index may provide a quantitative measure of whether a greater portion of easy items were rated too low or hard items were rated too high. Because the two standard deviation ratio indices captured similar trends and the second index can be applied with a wider range of IRT models, the second index may have greater utility in practical settings.

بخش اصلی

و تمایل انگیزشی برای موفقیت تحصیلی ارائه کرده‌اند ولی باز هم نمی‌توان به سهم و نقش جدید ذهن‌آگاهی بر میزان نگرش مثبت در مدرسه دست یافت.

قمری و حسینی (۱۳۹۶)؛ متین راد (۱۳۸۹) و بیرامی و عبدی (۱۳۸۸) هر سه در پژوهش‌های خود به نقش آموزش ذهن‌آگاهی بر کاهش اضطراب در محیط‌های آموزشگاهی اشاره کردند و با استناد به پژوهش‌های خارج از کشور مانند پژوهش بچمن، هاتچینز و پترسون[1] (۲۰۰۸) نیز می‌توان به این نتیجه رسید که در زمینهٔ تأثیر ذهن‌آگاهی بر اضطراب امتحان یافته‌های نسبتاً معتبری وجود دارد. پژوهش‌های بیرامی، موحدی، علیزاده گورادل (۱۳۹۴)؛ جشویند[2] و همکاران (۲۰۱۲)؛ آذرگون و کجباف (۱۳۸۹) و بیگل و براون[3] (۲۰۱۴) نیز آموزش ذهن آگاهی را عاملِ مهمی در جهت تقویت دیدگاهِ دانش‌آموزان نسبت به مدرسه و نسبت به معلم و همسالان می‌دانند و دریافتند که دانش‌آموزانی که با این شیوه آموزش دیده‌اند، بهتر از هم‌کلاسی‌های خود، توان سازگاری با محیط و همسالان را دارند. با این وجود، تفاوت‌های قابل توجهی که بین میزان اثربخشی پژوهش‌های آن‌ها وجود دارد، باعث شده است تا مسئلهٔ پژوهش با [illegible] حساسایی سهم و تأثیر آموزش ذهن‌آگاهی بر کاهش اضطراب امتحان [illegible] آن و افزایش نگرش نسبت به مدرسه و مؤلفه‌های آن، تدوین [illegible]

روش

پژوهش حاضر از نظر هدف کاربردی و از لحاظ روش، آزمایشی با طرح پیش آزمون–پس آزمون با گروه کنترل بود. جامعه آماری پژوهش حاضر را کلیهٔ دانش آموزانِ پایهٔ یازدهم شهر تهران در سال تحصیلی ۹۷-۱۳۹۶، تشکیل می‌دهند. به‌منظور انتخابِ نمونه از جامعه از روشِ نمونه‌گیری خوشه‌ای چندمرحله‌ای استفاده شد. به این صورت که به دلیل گستردگی شهر تهران، ابتدا شهر تهران به پنج منطقه جغرافیایی (شرق، غرب، شمال، جنوب و مرکز) تقسیم شد که از بین این خوشه‌ها، یک منطقه (منطقه شمال) به‌تصادف انتخاب شد و سپس از بین مناطق شمال تهران، دو منطقهٔ ۱ و ۳ به شیوه تصادفی انتخاب شدند. از بین مدارس مناطق ۱ و ۳، ۶ مدرسهٔ دخترانه و ۶ مدرسهٔ پسرانه از هر منطقه (مجموعاً ۱۲ مدرسه) به‌طور تصادفی

1. Beachemin, Hutchins & Patterson
2. Geschwind
3. Biegle & Brown

is improbable that raters would exhibit perfect fit, making the index overly sensitive as the number of items increases (Hurtz & Jones, 2009). Second, the index assumes statistical independence of ratings across items, which is often not the case. The index also looks at deviations across raters conceptualizing different cut scores. The index may yield high values due to individual cut score variation even if panelists exhibit reasonable fit to their own cut scores. We do not consider this index further in this article due to these challenges and because Hurtz and Jones (2009) did not recommend using the index.

An alternative index that uses squared deviations is the root mean squared difference (RMSD) between a panelist's item ratings and the expected probabilities of success:

$$\text{RMSD} = \sqrt{\sum_{i=1}^{n} [p_i - P_i(\theta)]^2 / n}, \tag{4}$$

where n is the number of items. The index is close to 0 when the item ratings and expected probabilities of success are close. The RMSD is a rater-level fit index, which allows it to be directly compared with the correlation-based index. The RMSD is not specifically designed to detect whether panelists are regressing their ratings for hard and easy items. We include this index in this study to see how an index based on squared deviations works to detect these effects.

Methods Based on Absolute Deviations

There are several indices based on absolute deviations in the research literature. Van der Linden (1982) suggested the most commonly used indices, the first of which looks at the mean absolute deviation between item ratings and the expected probabilities of success. The index is

$$E_{Ir} = \sum_{i=1}^{n} |p_i - P_i(\theta)| / n. \tag{5}$$

Van der Linden (1982) indicated that Equation 5 may be problematic because $P_i(\theta)$ is not equal to 0 or 1, meaning that the maximum possible deviation changes across items. He suggested the following transformation of Equation 5 that accounts for the maximum possible deviations:

$$C_{Ir} = \frac{E_{I,\max} - E_{Ir}}{E_{I,\max}}, \tag{6}$$

where

$$E_{I,\max} = \text{mean}\left(E_{i,\max}\right), \tag{7}$$

and

$$E_{i,\max} = \max\left\{P_i(\theta), 1 - P_i(\theta)\right\}. \tag{8}$$

$E_{I,\max}$ is the maximum possible error of specification. The ideal values for Equations 5 and 6 are 0 and 1, respectively. Similar to squared deviation indices, it is

زیر بخش یا بخش فرعی

روابط ریاضی

as evidence that ratings contain less error and are more consistent. There are two challenges with using such a correlation to investigate whether panelists may be regressing their ratings in toward the middle of the probability scale. First, p-values focus on the performance of all examinees. Angoff standard setting, however, focuses on minimally competent examinee performance. Hence, the correlation is not focused on the target population (Goodwin, 1999). While this is a conceptual challenge with these correlations, not focusing on minimally competent examinees in practice often only has a large impact when IRT models with discrimination parameters are used. In these cases, item ordering based on p-values is not necessarily the same as the item ordering at various ability levels, leading to somewhat different correlations if one were to compute a correlation based on the performance of all examinees versus one focused on minimally competent examinees. Second, this correlation is a measure of linear association between p-values and item ratings. Linearly transforming ratings so that ratings for hard items are too high and ratings for easy items are too low will not change this correlation. Such linear transformations are often what happen when panelists regress their item ratings in toward the middle of the probability scale.

The second type of correlation focuses on the performance of minimally competent examinees by using an IRT model. In this case, one computes the expected probability of success on the items using an IRT model and correlates this with the panelist's ratings (B. E. Clauser et al., 2002; B. E. Clauser et al., 2013; J. C. Clauser, Clauser, & Hambleton, 2014; Goodwin, 1999; Kane, 1994). For example, when using the Rasch (1960) model, one computes the expected probability of success as

$$P_i(\theta) = \{1 + \exp[-(\theta - b_i)]\}^{-1}, \tag{1}$$

where θ is a rater's estimated cut score on the underlying IRT ability scale and b_i is the difficulty of item i. Then, one simply computes the correlation as

$$r_{P(\theta)p} = cor\left[P_i(\theta), p_i\right], \tag{2}$$

where p_i is the Angoff rating for item i. High correlations are interpreted as representing ratings with less error. Equation 2 is a measure of linear association that also cannot detect a linear transformation of item ratings, meaning that Equation 2 will not be sensitive to the phenomenon we outline in this article. We include Equation 2 as a comparison index in this study.

Methods Based on Squared Deviations

Kane (1987) proposed evaluating Angoff ratings by computing a chi-square statistic using squared deviations as an aggregate measure of rater fit. Kane's index is

$$\sum_{i=1}^{n} Z^2_{iR} = \sum_{i=1}^{n} \left[\frac{\bar{p}_{iR} - P_i(\theta^*)}{\sigma_i(p_{iR})}\right]^2, \tag{3}$$

where $\bar{p}_{iR}$ is the mean item rating for item i across raters, $P_i(\theta^*)$ is the expected probability of success on item i at the estimated group cut score θ^*, and $\sigma_i(p_{iR})$ is the standard error of the item ratings. There are several issues with using Kane's index. First, it measures whether a group of raters perfectly fit a group cut score. It

عنوان جدول

Table 1

Value of Rater Indices Over Replications for θ = 0

		No Regression Effect (Baseline)						.6 × *b* + .1 Regression Effect						.3 × *b* + .1 Regression Effect					
		$r \approx .30$		$r \approx .60$		$r \approx .90$		$r \approx .30$		$r \approx .60$		$r \approx .90$		$r \approx .30$		$r \approx .60$		$r \approx .90$	
Index	Index Ideal	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
SD_b	1	1.00	.04	1.00	.04	1.00	.03	1.67	.08	1.68	.07	1.67	.05	3.35	.15	3.35	.15	3.34	.10
SD_p	1	1.01	.04	1.01	.04	1.00	.02	1.54	.07	1.54	.06	1.54	.04	2.94	.13	2.95	.13	2.94	.09
RMSD	0	.23	.01	.18	.01	.09	.00	.20	.01	.16	.01	.10	.00	.19	.00	.17	.00	.14	.00
E_{Ir}	0	.19	.01	.14	.01	.07	.00	.16	.01	.13	.01	.08	.00	.16	.00	.14	.00	.12	.00
C_{Ir}	1	.72	.01	.79	.01	.89	.01	.75	.01	.80	.01	.87	.00	.77	.01	.79	.01	.82	.00
$PWFI_{.10}$	1	.33	.03	.43	.03	.74	.03	.36	.03	.45	.03	.64	.03	.36	.02	.40	.02	.46	.02
$PWFI_{.05}$	1	.17	.03	.23	.03	.44	.04	.19	.03	.23	.03	.35	.03	.18	.02	.21	.02	.23	.02
Rating Balance	0	−.01	.03	−.02	.03	−.03	.04	−.03	.03	−.05	.03	−.07	.02	−.02	.02	−.02	.02	−.03	.01
cor	1	.30	.06	.60	.04	.90	.01	.30	.06	.60	.04	.90	.01	.30	.06	.60	.04	.90	.01

Note. SD_b is the ratio of the standard deviation of examinee item difficulty estimates over the standard deviation of the panelist item difficulty estimates in Equation 10, SD_p is the ratio of the standard deviation of expected probability of successes on the items over the standard deviation of the panelist item ratings in Equation 11, RMSD is the root mean square difference index in Equation 4, E_{Ir} is van der Linden's index from Equation 5, C_{Ir} is van der Linden's index from Equation 6, $PWFI_{.10}$ is the proportion of item ratings within a fixed interval of .10, $PWFI_{.05}$ is the proportion of items rating within a fixed interval of .05, Rating Balance is the rating balance index, and cor is the correlation between the item rating and the expected probability of success in Equation 2.

یادداشت جدول

may be systematically thinking that hard items are easier and easy items are more difficult than the values predicted by IRT models in Angoff standard settings. Research has shown such effects to be common across educational, psychological, and credentialing testing contexts, so being aware of and measuring these effects is important to having quality standard-setting data. Using the scatterplot, standard deviation ratio indices and other quantitative indices together will help practitioners to further understand the quality of Angoff ratings and improve standard-setting processes.

Acknowledgments

The authors would like to thank Dan Anderson for comments and suggestions he provided on an earlier version of this article. This article was written when the first author worked at the American Registry of Radiologic Technologists. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and are not necessarily the official position of the American Registry of Radiologic Technologists or Renaissance.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508–600). Washington, DC: American Council on Education.

Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education*, *17*, 59–88.

Clauser, B. E., Mee, J., Baldwin, S. G., Margolis, M. J., & Dillon, G. F. (2009). Judges' use of examinee performance data in an Angoff standard-setting exercise for a medical licensing examination: An experimental study. *Journal of Educational Measurement*, *46*, 390–407.

Clauser, B. E., Mee, J., & Margolis, M. J. (2013). The effect of data format on integration of performance data in an Angoff standard-setting exercise for a medical licensing examination: An experimental study. *International Journal of Testing*, *13*, 65–85.

Clauser, B. E., Swanson, D. B., & Harik, P. (2002). A multivariate generalizability analysis of the impact of training and examinee performance information on judgments made in an Angoff-style standard-setting procedure. *Journal of Educational Measurement*, *39*, 269–290.

Clauser, J. C., Clauser, B. E., & Hambleton, R. K. (2014). Increasing the validity of Angoff standards through analysis of judge-level internal consistency. *Applied Measurement in Education*, *27*, 19–30.

Engelhard, G. (2009). Evaluating the judgments of standard-setting panelists using Rasch measurement theory. In E. V. Smith, Jr., & G. E. Stone (Eds.), *Criterion referenced testing: Practice analysis to score reporting using Rasch measurement models* (pp. 312–346). Maple Grove, MN: JAM Press.

Goodwin, L. D. (1999). Relations between observed item difficulty levels and Angoff minimum passing levels for a group of borderline examinees. *Applied Measurement in Education*, *12*, 13–28.

Humphry, S., Heldsinger, S., & Andrich, D. (2014). Requiring a consistent unit of scale between the response of students and judges in standard setting. *Applied Measurement in Education*, *27*, 1–18.

قدردانی

مراجع

عنوان جدول

جدول ٦. تحلیل واریانس برای بررسی همگنی شیب‌های رگرسیون در متغیر اضطراب امتحان

متغیر	منبع تغییرات	مجموع مجذورات	df	میانگین مجذورات	F	سطح معناداری
اضطراب امتحان	گروه×پیش‌آزمون اضطراب امتحان	۱۴/۱۴۳	۱ و ۸۴	۱۴/۱۴۳	۰/۸۴۰	۰/۳۶۲

یکی دیگر از مفروضه‌های تحلیل کوواریانس تک متغیری و چند متغیری، مفروضهٔ همگنی شیب‌های رگرسیون[1] است. اطلاعات جدول ۶ نشان می‌دهد که سطح معناداری مقدار F به‌دست‌آمده بالاتر از ۰/۰۵ است ($F_{(1,84)}=0/840, P_s=0/362$)؛ بنابراین شیب‌های رگرسیون پیش‌آزمون و پس‌آزمون نمرات اضطراب امتحان در دو گروه آموزش ذهن آگاهی و گروه کنترل تفاوت معناداری نداشته و فرض همگنی شیب‌های رگرسیون تأیید می‌شود.

عنوان جدول

با توجه به برقراری مفروضه‌های تحلیل کوواریانس تک متغیری، انجام این آزمون جهت تعیین اثربخشی ذهن‌آگاهی بر اضطراب امتحان کل مجاز است.

جدول ۷. خلاصه تحلیل کوواریانس تک متغیری برای اثربخشی آموزش ذهن آگاهی بر اضطراب امتحان کل

گروه	پیش‌آزمون		پس‌آزمون		مقدار F	درجه آزادی	سطح معناداری	مجذور اتا
	میانگین	انحراف معیار	میانگین	انحراف معیار				
ذهن‌آگاهی	۲۵/۶۱	۳/۸۸	۴۱/۰۲	۳/۷۱	۱۹۳/۵۷۱	۸۵ و ۱	۰/۰۰۱**	۰/۶۹۵
کنترل	۲۴/۷۰	۳/۵۹	۲۸/۵۰	۴/۶۸				

**P<۰/۰۱, n = ۸۸

نتایج تحلیل کوواریانس نشان می‌دهد که بین میانگین نمرات اضطراب امتحان گروه آموزش ذهن‌آگاهی و گروه کنترل تفاوت معناداری وجود دارد ($F_{(85,1)}=193/571 P_s<0/01$). میزان تأثیر ۰/۶۹۵ بود. با توجه به تفاوت افزایش میانگین‌ها در مرحله پس‌آزمون نسبت به مرحله پیش‌آزمون که در گروه ذهن آگاهی این افزایش میانگین‌ها بیشتر است و با توجه به سطح معناداری به‌دست‌آمده، بنابراین می‌توان گفت که آموزش ذهن آگاهی بر کاهش اضطراب امتحان دانش‌آموزان تأثیر معناداری دارد. مجدداً

1. homogeneity of regression

مراجع

قوی‌تری نسبت به مدرسه منجر می‌شود (تاپولی و همکاران، ۲۰۰۵). در نتیجه می‌توان این گونه استدلال کرد که مداخلهٔ ذهن‌آگاهی به دلیل ایجاد نظارت فراشناختی و آموزش تکنیک‌های بودن در لحظهٔ حال به دانش آموزان کمک می‌کند تا بر پردازش افکار ناکارآمد و نگرش‌های منفی خود غلبه کنند و با آگاهی بیشتری با آن‌ها برخورد کنند، لذا نگرش منفی آن‌ها نسبت به مدرسه به میزان زیادی کاهش یابد. به‌طورکلی نتایج به‌دست‌آمده مؤید این مطلب است که آموزش ذهن‌آگاهی به دانش‌آموزان کمک کرد که درک درستی از هیجانات خود داشته باشند، آن‌ها را به‌طور کامل تجربه کنند و از رویکردهای ذهنی خودآگاه شوند، در واقع این تجربهٔ کامل هیجان‌ها و این خودآگاهی نسبت به هیجانات موجب شد تا هنگام اضطراب و یا هنگام هجوم نگرش‌ها و افکار منفی در موقعیت‌های مختلف، بتوانند رویکرد ذهنی خود را از سبکی به سبک دیگر تغییر دهند و هیجان حاصل از این موقعیت‌ها را کاهش یا افزایش داده و با آن مقابله کنند.

محدود بودن جامعه پژوهش به یک پایهٔ تحصیلی (پایهٔ دوازدهم) از محدودیت‌های این پژوهش بود، لذا پیشنهاد می‌شود پژوهش‌هایی بر روی دانش‌آموزان سایر مقاطع تحصیلی صورت پذیرد. همچنین از محدودیت‌های دیگر این پژوهش، عدم استفاده از سایر مداخله‌های روان‌شناختی به‌منظور مقایسه است، در این راستا پیشنهاد می‌شود از سایر شیوه‌های آموزشی در کاهش اضطراب امتحان و افزایش نگرش به مدرسه و مقایسه آن‌ها با نتایج این پژوهش استفاده شود و درنهایت با توجه به اثربخشی آموزش ذهن‌آگاهی بر متغیرهای مذکور، پیشنهاد می‌شود که این برنامه در محیط‌های آموزشی توسط روان شناسان و مشاوران برای دانش آموزان اجرا شود.

منابع

آذرگون، حسن و کجباف، محمدباقر. (۱۳۸۹). اثربخشی آموزش ذهن‌آگاهی بر کاهش نگرش ناکارآمد و افکار خودآیند دانشجویان افسرده دانشگاه اصفهان. *مجلهٔ روان‌شناسی*، ۱۴ (۱)، ۲۶-۲۱.

باعزت، فرشته؛ صادقی، منصوره سادات؛ ایزدی فرد، راضیه و روبن زاده، شرمین. (۱۳۹۱). اعتباریابی و رواسازی آزمون اضطراب امتحان فریدین. *مجلهٔ مطالعات روان‌شناختی*، ۸ (۱)، ۶۶-۵۱.

Lawson, D. J. (2006). Test anxiety: A test of attentional bias.

Lufi, D., Okasha, S., & Cohen, A. (2004). Test anxiety and its effect on the personality of students with learning disabilities. *Learning Disability Quarterly*, *27*(3), 176-184.

McCoach, D. B., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement*, *63*(3), 414-429.

Miller, D. G. (2016). The Impact of Formal Extracurricular Activities on Satisfaction and Attitudes-toward-School among At-Risk Adolescents.

Napoli, M., Krech, P. R., & Holley, L. C. (2005). Mindfulness training for elementary school students: The attention academy. *Journal of applied school psychology*, *21*(1), 99-125.

Pena, M., & Losada, L. (2017). Test anxiety in spanish adolescents: examining the role of emotional attention, and ruminative self-focus and regulation. *Frontiers in psychology*, *8*, 1423.

Perri, S. (2017). The Impact of a Mindfulness-Based Intervention on Elementary School Students Anxiety Level.

Steele, C. M. (2011). *Whistling Vivaldi: How stereotypes affect us and what we can do (issues of our time)*. New York, NY: WW Norton.

Stöber, J., & Esser, K. B. (2001). Test anxiety and metamemory: general preference for external over internal information storage. *Personality and Individual Differences*, *30*(5), 775-781.

Takamiya, Y., & Tsuchiya, S. (2018). The self-care education for healthcare professional students using mindfulness. *International Journal of Whole Person Care*, *5*(1).

Teens, S. (2010). Assessing the efficacy of an adapted in-class mindfulness-based training program for school-age children: A pilot study.

Trifoni, A., & Shahini, M. (2011). How does exam anxiety affect the performance of university students. *Mediterranean Journal of Social Sciences*, *2*(2), 93-100.

ASSIGNMENT