

Research Article

Open Access

Application of the Angoff Method for Assessing Item Appropriateness and Setting Cut-off Scores during the Pre-Clinical Instrument Development Phase

Kimberly A Williams^{1*}, Shadi Kanan¹ and Lynette Phillips²

¹College of Nursing, Kent State University, Kent, OH, USA

²College of Public Health, Kent State University, Kent, OH, USA

Article Info

***Corresponding author:**

Kimberly A Williams

College of Nursing
Kent State University
P.O. Box 5190 Kent, OH 4424
330-672-8823
USA
E-mail: kwill190@kent.edu

Received: May 25, 2018

Accepted: June 19, 2018

Published: June 25, 2018

Citation: Williams KA, Kanan S, Phillips L. Application of the Angoff Method for Assessing Item Appropriateness and Setting Cut-off Scores during the Pre-Clinical Instrument Development Phase. *Madridge J Nurs.* 2018; 3(1): 113-117.
doi: 10.18689/mjn-1000120

Copyright: © 2018 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Published by Madridge Publishers

Abstract

Setting cut-off scoring that indicates positive findings for sensitive behaviors is important but can be difficult in the pre-clinical testing phase of instrument development. This article discusses the use of the Angoff Method to evaluate the appropriateness of items and set cut-off scores for a non-suicidal self-injurious behavior screening tool. Findings from this study demonstrated that the agreement for item appropriateness for assessing non-suicidal self-injury was 97.2%. Cut-off were calculated by using the cumulative probabilities, as percentages, then summed to provide an average cut-off score for low-, moderate-, and high-risk. After completion of the study, this method was found to be a simple approach for setting cut-off scores during pre-clinical testing. Limitations did surface indicating a need for further studies using the Angoff Method for setting pre-clinical cut-off scores in instrumentation.

Keywords: Angoff method; Pre-clinical instrument; Setting cut-off scoring; Non-suicidal self injury.

Introduction

Instrumentation is often used in nursing research to help develop reliable instruments. The steps involved in instrumentation include searching the literature for a pre-existing concept or development of a concept, identifying content items, and then validating these items for appropriateness through experts. After validation, missing items are added to the instrument and assessed for reading level and clarity. Once these four steps have been completed, pilot testing is then performed to assess additional characteristics and internal consistency. Once completed, the instrument is then tested in a large population simultaneously with tools measuring a similar construct.

The developed items of the instrument in the initial stage include decisions of the type of format used to indicate an answer for each item. Formats used include Likert or Likert like scales, a simple check box, or yes/no answer options. In addition, some screening instruments may need to conclude that a particular person does or does not have fit into a specific group such as low- moderate- and high-risk. To categorize a person into a specific group, cut-off scores are needed to set the minimum and maximum score for each group. In this situation, structure assessment is performed by factor analysis; cut-off scores are informally set, and then analyzed later for accuracy using a large sample of study participants. However, there may be circumstances when researchers need to set formal cut-offs in the pre-clinical phase prior to studying the instrument in a pilot or a large population. This is especially true when screening for sensitive issues such as mental disorders, suicidal behavior and non-suicidal self-injury. Therefore, researchers developing clinical screening instruments may want to

utilize a different approach for determining cut-off scores when placing a person in a low-, moderate- or high-risk category. This article offers a unique use of the Angoff Method to set cut-offs during the pre-clinical stage.

The Angoff Method was first introduced in 1971 [1], [2] and utilizes expert opinions for assessing item appropriateness and setting cut-off scores for standardized test [3], [4]. Initially, experts in the content area imagine a borderline case (person), conceptualize this borderline case, examine each item for appropriateness, and assess the probability that this conceptualized case would answer each item correctly (or in a specific manner). Cut-offs scores for passing or positive cases are set when there is a high degree of agreement among these judges. Although it has been used over the past 45 years in the educational setting to set minimum passing scores for standardized testing, researchers have suggested the Angoff Method to set cut-off scores for health screening instruments. According to Goodwin [5], this method may be used in clinical situations in which cut-off scores are needed for mental screening measures to distinguish areas such as "normal-no problem identified," "borderline-possible problem identified," and "atypical-definite problem identified." Considering this suggestion, we conducted a literature search in 2017 and found that the Angoff Method had not been used to set cut-off scores for health assessment instrumentation such as screening tools. Therefore, we sought to apply this method for setting pre-clinical cut-off scores for a clinician administered screening tool for non-suicidal self-injury.

Method

Angoff Method

The Angoff Method related to setting minimum performance on academic examinations consists of nine steps [2]. The initial steps involve (1) identification of seven to ten experts or judges, (2) require a minimum of 50% agreement for inclusion of an item, (3) ask experts or judges to evaluate the grading key correctness, and (4) identify that each item is related to the area being tested. Next, (5) experts are asked to describe what would be the consequence of their future performance for the person taking the test if they did not know the answer to the item. The experts are asked (6) how easy an item could be looked up within their daily function and (7) to differentiate if the item's cut-off is higher than the minimum competency of the test taker. Finally, experts are asked to (8) conceptualize the minimal acceptable person and the probability that they would answer the question correctly from a probability of 0 to 100%. The last step (9) is that the items' cumulative probabilities, as percentages, are then summed to provide an average cut-off score for the examination.

Because our study involved the use of the Angoff Method to evaluate items for appropriateness and relevance and to determine cut-off scores for low-, moderate- and high-risk categories of a clinician-administered screening tool for risk of non-suicidal self-injury (NSSI), we utilized seven of the nine steps. Given this focus, we did not ask each expert to ensure the grading key was correct, or if an item could be easily

looked up. We posed the following questions: Is each item for NSSI behavior appropriate for screening? What percentage will expert judges assign to each characteristic that exists for non-suicidal self-injurers in their practice? What is the cut-off score for low-, moderate-, and high-risk for the clinician-administered screening tool?

Item Development

Items describing the characteristics of NSSI for the Angoff Method were taken from two investigator-developed screening tools previously designed from a literature review related to non-suicidal self-injury [6], [7], [8], [9]. Because items 1 through 27 were from the self-administered screening tool in a Likert scale format, only items 4, 7 and 8 pertaining to gender and age ranges were included in this analysis. Of the 51 items, 28 through 75 were items specifically from the binomial clinician-administered screening tool, producing a total score that resulted in an estimated risk value. Although there were 51 items, the highest score achievable on the screening tool was 41. The integration of two items of age ranges (13 to 15 and 16 to 19) from the self-administered screening tool was the cause for this discrepancy. The age item of the clinician administered screening tool included one binomial item of 12 and older which was not an item on the form completed by judges in this study.

The Angoff matrix incorporated age groupings commonly used in demographic surveys such as those used in the U.S. Census surveys. There were also three sections comprising the clinician-administered screening tool that included items which had scorings equaling less than the number of items included in that section. For example, one section asked clinicians to indicate the appearance of any scars and wounds (6 items). The highest score for the entire section was four. Also, two points were given if there was one scar or wound and four points were given if there were more than one scar or wound. The items describing NSSI characteristics entered into the Angoff matrix so that the expert judge could easily respond (see table 1). The document was then sent to an instrument development specialist who suggested minor wording revisions. Following revisions, approval for the study was obtained from the university's institutional review board.

Table 1. Example of Angoff Averages of Percentage

1. Age range: 10-12 years	<input type="checkbox"/> No	Injury	None/Low	Moderate	High/Extensive
Comments:	<input checked="" type="checkbox"/> Yes	Percent	75%	20%	5%

Note: numbers refer to the percentage that the item has been seen in persons in the judge's client population with NSSI at no/low, moderate and high risk.

Judge Selection and Orientation

To obtain a sufficient number of expert judges to evaluate the items, a flyer was distributed via email to a local community mental health facility. The flyer asked providers with experience caring for patients with NSSI behaviors to participate in a study to develop a screening tool for these behaviors. Ten providers returned consent forms to participate and received a nine-page packet that included a reviewer's profile, instructions for rating NSSI criteria, and an Angoff

matrix consisting of 75 NSSI items for evaluation. Next, expert judges were asked to provide their discipline, credentials, highest degree earned, total years in psychiatric practice and total years treating patients with NSSI behaviors. Instructions included in the packet asked participants to perform two different evaluations of each item. First, the expert judges were asked to decide whether each item was applicable in any way to NSSI, with items considered unrelated to NSSI identified. Next, the judges were asked to estimate the percentage of none/low, moderate, or high/extensive items seen in non-suicidal self-injurers in their practice for each item deemed appropriate for NSSI. For example, a rating of 50 would mean that judges had noted a characteristic, behavior, symptom, or risk factor in about 50% [2], of those with NSSI.

Results

A variety of psychiatric providers (n=10) participated as expert judges in this study. These included five psychiatric nurse practitioners, two psychiatrists, and three social workers. Education levels for participants included master's degrees (seven) and doctoral degrees (three). All held positions working with a psychiatric population and had experience with treating NSSI behaviors ranging from 1.75 to 20 years, with an average of 8.4 years.

Item Analysis

Item analysis was examined by evaluating the percentage of agreement of appropriateness for each item, or the agreement among expert judges who indicated "yes" or "no" for each item. Items of 50% or more for appropriateness were considered. The total percentage for items (across the three classifications of injury groupings) was between 80% and 100% for the risk factor minimally seen in the NSSI patient population for whom they provided care (see table 2).

Table 2. Angoff Matrix Data for All 10 Participants

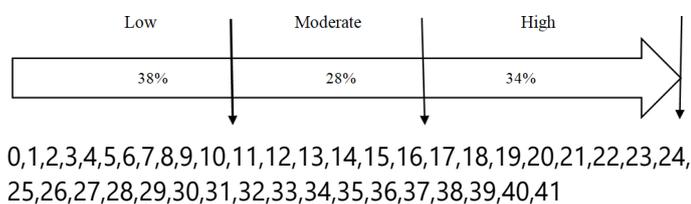
Angoff Results for Questions for the Clinician-Administered Screening Tool	Items Appropriate Yes/No	Injury: None/Low	Injury: Medium	Injury: High/Extensive
4. Is Female	9/0*	0.405	0.323	0.273
7. Age 13-15 years	10/0	0.563	0.258	0.178
8. Ages 16-19 years	10/0	0.475	0.288	0.202
28. Has only one scar or wound likely caused by NSSI	9/1	0.386	0.378	0.27
29. Has two or more scars or wounds likely caused by NSSI	10/0	0.27	0.312	0.418
30. Has scars or wounds on arm likely caused by NSSI	10/0	0.31	0.30	0.39
31. Has scars or wounds on leg likely caused by NSSI	10/0	0.297	0.32	0.383
32. Has scars or wounds on abdomen likely caused by NSSI	10/0	0.355	0.273	0.372
33. Has scars or wounds on chest likely caused by NSSI	9/1	0.39	0.24	0.37
34. Has scars or wounds on another area likely caused by NSSI	10/0	0.475	0.227	0.298
35. Scars or wounds appear to be burn	10/0	0.434	0.288	0.278
36. Scars or wounds appear to be linear	10/0	0.248	0.268	0.483
37. Scars or wounds appear to be jagged	10/0	0.47	0.246	0.284
38. Scars or wounds appear to be round and smooth	10/0	0.505	0.238	0.247
39. Scars or wounds appear to be ____ (not listed above).	4/4**	0.50	0.32	0.18

40. Description of how the injuries/scar occurred fits the appearance of the scars or wounds	8/1*	0.474	0.309	0.216
41. Description of how the injury/scar occurred does not fit the appearance of the scars or wounds	9/1	0.433	0.268	0.309
42. Wears inappropriate clothing for the setting or seasons	10/1	0.33	0.263	0.407
43. Expresses excessively uncomfortable/refuses to change normal situations: i.e. health exams or sport activities	8/1*	0.478	0.253	0.269
44. Style of clothing is unusual and <i>cannot</i> be attributed to culture or religion	9/1	0.448	0.264	0.288
45. Appears to have extreme moodiness (i.e. fights at school)	10/0	0.393	0.277	0.324
46. Appears to have inappropriate expressions of anger (aggression)	10/0	0.418	0.297	0.253
47. Appears to have poor self-esteem	10/0	0.22	0.255	0.535
48. Appears to have poor impulse control	10/0	0.27	0.265	0.465
49. Appears sad or tearful	10/0	0.313	0.308	0.378
50. Expresses feelings of anxiety or stress	9/0*	0.313	0.313	0.373
51. Appears anxious or stressed	10/0	0.37	0.28	0.345
52. Appears to be an underachiever	9/1	0.531	0.237	0.231
53. Appears to be an overachiever	9/1	0.378	0.267	0.344
54. Expresses feelings of self-disappointment	10/0	0.366	0.244	0.385
55. Appears low or down	10/0	0.325	0.275	0.395
56. Is unable to express positive aspects of his or her life	9/1	0.344	0.259	0.381
57. Has tattoos/body piercing/body gauges (please provide more information on location, number, style if possible)	10/0	0.467	0.261	0.278
58. Patient has an eating disorder, either excessive or restrictive eating	10/0	0.34	0.30	0.36
59. Patient has plucking of hair on scalp, eyelashes, eyebrows or genital areas	9/1	0.435	0.26	0.305
60. Patient uses of alcohol &/or substances regularly	10/0	0.493	0.223	0.283
61. Has difficulty adjusting to new situations	9/1	0.428	0.283	0.305
62. Has a history or has recently been exposed to a traumatic event	10/0	0.305	0.29	0.405
63. Is atypically developed or shows evidence of developmental delay	9/1	0.567	0.239	0.144
64. Has a history of suicide attempt(s)	10/0	0.33	0.272	0.398
65. Currently under the care of a psychiatrist or other therapist	10/0	0.268	0.293	0.438
66. Currently on psychiatric medications (specify)	10/0	0.267	0.308	0.428
67. Patient has a history of or is currently an abuse victim	10/0	0.35	0.225	0.425
68. Patient has a history of or is currently a sexual abuse victim	10/0	0.285	0.22	0.495
69. Patient has a history of legal trouble or is currently in juvenile incarceration	10/0	0.36	0.26	0.385
70. Family history/parental guardian history of substance use or abuse	10/0	0.315	0.295	0.39
71. Family violence/turmoil in the home	10/0	0.27	0.30	0.425
72. Feels the need to make good grades or have academic success	9/1	0.478	0.294	0.228
73. Experiences excessive peer pressure when compared to others of a similar age and situation	9/0*	0.375	0.29	0.335
74. Lives in an abusive environment (abuse directed at patient or others)	9/0*	0.333	0.326	0.34
75. Is a victim of bullying (physical, cyber or otherwise)	9/0*	0.293	0.353	0.353
Angoff Results	Appropriate	Low	Medium	High
Totals	Yes = 524 No=18	0.381	0.279	0.338

Note: that not all experts rated each item: * Missing one variable; ** Missing two variables

The overall agreement of 96.7% was achieved. The percentages for items 4, 7, 8, and 28 through 75 (see table 2) were averaged under each column of none/low, moderate, or high/extensive. Next, the percentages (see table 2) and the total score (41) for the screening tool was used to calculate cut-off scores. Low-risk was determined by calculating the score that would be in the lower 38% of the total score of 41 points from the clinician-administered tool. The low-risk score ranged between 1 and 16 points; therefore, if the total score of the clinician administered tool was 16 points or fewer, the person scored in the low-risk category. Moderate-risk was determined using the next 28% of the total, resulting scores of 17 to 27 points for those in the moderate-risk category. Lastly, high-risk was determined by using the remaining 34% of the total for a score of 28 to 41 points, indicating placement in the high-risk category (see Figure 1).

Figure 1. Cut-off Points



Discussion

The usual practice is to set cut-offs using a clinical population in the last step of instrumentation. When developing an instrument assessing sensitive behaviors and categorizing a person into groupings of low-, moderate-, and high-risk; setting pre-clinical cut-off scores before testing in a clinical population was needed. In our search, we were unsuccessful in locating a method previously used in health care screening tool development to apply. We decided to apply the Angoff Method, a well-researched method used by other disciplines such as academia, in a unique way. We did find this a useful approach. In our findings, we noted that the percentages of the total score were higher for low-risk (38%) than for high-risk cases (34%). This finding is reasonable, as you would expect to find a larger population of those who do not perform NSSI behavior than those who do perform NSSI behavior. For example, some studies have shown prevalence rates among young adults are 14-17% and among adolescents is 13-23% [10-12].

There were limitations to this study. The clinician administered tool indicated age 12 and older as an item. However, we were not able to include the data for age 12 in this analysis as it was collected in the group data for ages 10 to 12. Additionally, we had 10 psychiatric providers participate as expert judges; a number that is considered appropriate for this study [2, 5]. Had more psychiatric providers participated, they may have provided different results. Also, we assumed that providers were aware of levels of risk related to their current practice and working with persons who perform NSSI. Therefore, we did not give a specific definition for low-, moderate-, and high-risk. We noted a middling result with this analysis, and we speculate that some participants with less experience in the practice setting and not completely understanding low-, moderate-, and high-risk may have influenced this finding. For example, participants new to

the field may have had "difficulty conceptualizing" the person with NSSI behavior [5], may not have considered the criteria fully for risk levels, and their judgment may not have been as accurate as those with more experience. Thus, judges with less experience may not have adequately categorized a low-, moderate-, or high-risk patient.

The fact that participants completed the Angoff matrix form at their own pace may have affected participants' ability to adequately categorize items. There seemed to be a lack of clear designation for each item as items were not clearly placed under low-, moderate-, or high-risk. While the Angoff matrix included very detailed instructions, providers may have experienced difficulty completing the form and did not shared this information with the researcher. Future studies using the Angoff Method for cut-off scores should consider a focus group session, allowing participants to discuss each item as a group and allowing researchers to answer participant questions regarding completion of the Angoff form. Additionally, new researchers attempting to "use the Angoff Method are often given empirical item P-values, and may not know how to use" the data generated [12]. Therefore, researchers may want to consider additional training or consult with a specialist familiar with the Angoff Method. Lastly, this is the first known study using this method to establish cut-off scores for a screening tool. Thus, we have not yet studied the results in comparison to the traditional methods for setting cut-off scores.

Nursing Implications

We found the Angoff Method to be a potentially useful method for evaluating the appropriateness and relevance of an instrument's items. In addition, this study found the Angoff Method provides a rigorous method of setting cut-off scores for instrument development during the pre-clinical development phase. However, we recommend future research studies of instrumentation be conducted using this method and that they include comparison of it to other reliable methods for setting cut-off scores. In doing so, future studies may validate this method as a means to establish cut-off scores for instrument development in pre-clinical stage.

Conflicts of Interest

The authors have no conflict of interest.

Funding

This study received funding from the University of South Alabama College of Nursing Dean's grant.

References

1. Angoff WH. Scales, norms and equivalent scores. In Thorndike RL, ed. *Educational Measurement*. 2nd ed. Washington, DC: American Council on Education. 1971; 508-600.
2. Ricker KL. Setting cut-scores: A critical review of the Angoff and modified Angoff methods. *AJER*. 2006; 52(1): 53-64.
3. Berk RA. A consumer's guide to setting performance standards on criterion referenced tests. *Rev Edu Res*. 1986; 56(1): 137-172. doi: 10.3102/00346543056001137

4. Biddle RE. How to set cutoff scores for knowledge tests used in promotion, training, certification, and licensing. *Public Pers Manage*. 1993; 22(1): 63-79. doi: 10.1177/009102609302200105
5. Goodwin LD. Focus on quantitative methods: Determining cut-off scores. *Res Nurs Health*. 1996; 19(3): 249-256. doi: 10.1002/(SICI)1098-240X(199606)19:3<249
6. Williams KA, Phillips L, Chadwell J, Monsman H. Non-Suicidal Self-Injury Risk Screening in Primary Care Settings: Results of Two Instrument Development Studies. *Madridge Journal of Nursing*. 2018; 3(1): 92-100.
7. Williams KA, Bydalek K. Self-mutilation: The cutting truth. *American Nurse Today*. 2009; 4(8): 29-32.
8. Williams KA, Bydalek K. Adolescent self-mutilation: Diagnosis and treatment. *J Psychosoc Nurs Ment Health Serv*. 2007; 45(12): 19-23.
9. Williams KA, Yoo W, Rikabi KK. Understanding Non-Suicidal Self-Injury Assessment in the Primary Care Setting Usefulness of the SLICE Tool. *Journal of the Student National Medical Association*. 2013; 6: 21-22.
10. Polanco-Roman L, Jurska J, Quinones V, Miranda R. Brooding, reflection, and distraction: Relation to non-suicidal self-injury versus suicide attempts. *Archives of Suicide Research*. 2015; 19(3): 350-365. doi: 10.1080/13811118.2014.981623
11. Saraff PD, Trujillo N, Pepper CM. Functions, consequences, and frequency of non-suicidal self-injury. *Psychiatric Quarterly*. 2015; 86(3): 385-393. doi: 10.1007/s11126-015-9338-6
12. Skaggs G. Coming full circle in standard setting: A commentary on Wyse. *Measurement: Interdisciplinary Research and Perspectives*. 2013; 11(4): 195-197. doi: 10.1080/15366367.2013.857217