DOCUMENT RESUME

ED 400 299 TM 025 616

AUTHOR Rich, Charles E.; Johanson, George A.
TITLE An Item-Level Analysis of "None of the Above."
PUB DATE Apr 90
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Difficulty Level; *Item Analysis; *Multiple Choice Tests; *Test Construction; Test Reliability
IDENTIFIERS *None of the Above (Multiple Choice Tests)

ABSTRACT

Despite the existence of little empirical evidence for their effectiveness, many techniques have been suggested for writing multiple-choice items. The option "none of the above" (NA) has been widely used although a recent review of empirical studies of NA suggests that, while generally decreasing the difficulty index, NA also decreases discrimination and may decrease reliability. It is suggested that most of the studies of the effect of NA on these item parameters have been flawed by methodological inconsistencies and by a disregard for the finding that discrimination is restricted when corresponding item difficulties have been extremely high or low values. By examining the effects of NA on difficulty and discrimination indices in light of optimal difficulty for a 100-item test taken by 300 undergraduates, this study found that when following reasonable guidelines: (1) difficulty tended to approach the optimal level; (2) discrimination tended to increase; and (3) reliability was unaffected. (Contains 6 tables and 24 references.) (Author/SLD)

An Item-level Analysis of "None of the above"

Introduction

Over the years a number of studies have presented rules for writing multiple-choice tests (Ebel, 1951; Wesman, 1971; Haladyna & Downing, 1989), and yet these authors have pointed out that there is often little empirical evidence for these item-writing rules. It has been suggested that the choice of distractors in multiple-choice items is the most important aspect of item writing (Hopkins & Stanley, 1981; Weitzman & McNamara, 1946). However, any experienced item writer knows that it is often difficult to develop enough good distractors, and would welcome any valid technique that could simplify the process. One such technique often recommended is that of using "none of the above" (NA) (Stanley, 1964; Thorndike & Hagen, 1969; Ebel, 1979; Roid & Haladyna, 1982; Nitko, 1983; Mehrens & Lehman, 1984). Some of the advantages mentioned in these studies are that NA is an easy way to develop an extra option in items where options are hard to devise, a way to decrease the chances of guessing correctly, and a good replacement for a weak distractor.

The use of NA is not without controversy. In a recent review of the validity of item-writing rules, Haladyna & Downing (1989) summarize several empirical studies of the effects of NA on item and test parameters and interpret the results as generally discouraging the use of NA. These authors report that the overall results from six studies were that NA decreases the difficulty index (makes item harder) and decreases discrimination and reliability as well. Although Haladyna & Downing (1989) do call for more research on the NA option, the present authors anticipate that since these findings were presented as an overall review of the literature they may be taken as more conclusive than is justified. In the present paper it is suggested that questionable and/or differing methods in the reviewed studies preclude generalizing as to the effect of reasoned use of NA on item difficulty and discrimination. Item-writing guidelines specifically for the use of NA are suggested, and evidence is presented that NA can be used effectively to move item difficulty indices into a moderate, "optimal" range which may permit increases in item discrimination and test reliability. In stressing the importance of item discrimination, a norm-referenced perspective on measurement is assumed.

This paper is presented in three main sections: a) a review of methods used in studies reviewed by Haladyna & Downing (1989) and two other relevant studies, b) a discussion of what the present authors call for the sake of simplicity the "optimal difficulty approach" for investigating the effect of NA on item difficulty and discrimination, supported in part by a re-analysis of data from Wesman & Bennet (1946) and Tollefson (1987), and by c) results from two studies conducted by the present authors in which the optimal difficulty approach is used to investigate effects of NA on item discrimination and test reliability.

An Item-level Analysis of "None of the above"

Charles E. Rich & George A. Johanson
Ohio University

A paper presented at the annual meeting of the American Educational Research Association
April 1990
Boston, Massachusetts

Abstract

Despite little empirical evidence for their effectiveness, many techniques have been suggested for writing multiple-choice items. The option "none of the above" (NA) has been widely used although a recent review of empirical studies of NA suggested that, while generally decreasing the difficulty index, NA also decreases discrimination and may decrease reliability. In the present study it is suggested that most studies of the effect of NA on these item parameters have been flawed by methodological inconsistencies and by a disregard for the finding that discrimination is restricted when corresponding item difficulties have extremely high or low values. By examining the effects of NA on difficulty and discrimination indices in light of optimal difficulty, this study found that when following reasonable guidelines (1) difficulty tended to approach the optimal level, (2) discrimination tended to increase, and (3) KR-20 reliability was unaffected.

Methodology

Review of current techniques

For the present study the authors were able to locate five of the six studies cited by Haladyna & Downing (1989), which dealt with NA and item difficulty and discrimination (Wesman & Bennet, 1946; Hughes & Trimble, 1965; Dudycha & Carpenter, 1973; Mueller, 1975; Forsyth & Spratt, 1980) dealt with five issues which exist when using NA. One of the articles (Schmeiser & Whitney, 1975) could not be located. Two other studies of NA, item difficulty and discrimination (Tollefson & Tripp, 1986; Tollefson, 1987) were also reviewed.

Although the following six methodological issues are not exhaustive of issues regarding NA, it was around them that much inconsistency in method revolved in the reviewed studies. These issues are: a) how to select the distractor to be replaced by NA, b) how often to use NA as the correct option, c) what proportion of the test items can effectively include NA, d) whether NA should be used as the correct option on mathematics items requiring calculations, e) how to select items whose item parameters can be improved by using NA, and f) how to assess the effects of NA on item difficulty and discrimination. These issues are dealt with below.

Distractor selection. It is generally recommended that a weak distractor be replaced with a more attractive one (Wesman & Bennet, 1946; Dudycha & Carpenter, 1973). Accordingly, it would seem reasonable to subsitute NA for the weakest distractor, as was done by Dudycha & Carpenter (1973). They do not supply a rationale but common sense dictates that NA could be more attractive than a distractor which seems irrelevant. Wesman & Bennet (1946) suggest that the effectiveness of NA may be most dependent on the quality of the other options in the item. Those authors suggest also that to replace an effective distractor with NA may cause no effect or may be detrimental, while replacing a weak distractor may improve the item. A wide variety of methods were employed in the reviewed studies for selecting distractors to be replaced by NA: simply substituting NA for the fifth option of each item (Wesman & Bennet, 1946), random selection (Dudycha & Carpenter, 1973), adding a fifth option (NA) to 4-option items (Hughes & Trimble, 1965), and, curiously, substituting NA for the most frequently chosen distractor (Tollefson, 1987). The three remaining studies did not indicate their method of substitution.

Number of items using NA. If NA were used on every item of a test it is very likely that the credibility of NA would suffer, since examinees might view NA as filler used due to ignorance or laziness (Osterlind, 1990), and since few examinees would assume it could be the correct answer on every item. The authors suggest that at most it be used on the percentage of items which is the inverse of the number of options used on the items. For example, for tests utilizing 4-option items, a maximum of 25% of the items could include NA. The rationale is the same as that for using a balanced key to maintain the credibility of each option. Hence, the examinee should sense that if NA occurred in position D on 25 items of a 100-item test, it could theoretically be correct on each of those 25 items. On tests containing items with different numbers of options, the percentage could be based on finding the average number of options and then interpolating.

This percentage, based on the number of options, was greatly exceeded in most of the reviewed studies: 100% of the items (Wesman & Bennet,1946; Forsyth & Spratt, 1980; Tollefson, 1987); from 38% to 73% (Hughes & Trimble, 1965; Dudycha & Carpenter, 1973; Tollefson & Tripp, 1987). Mueller (1975) used NA on only about 16% of his items but also included "all of the above" and options like "both a and b are correct" so that from 28% to 60% of his items consisted of these "complex alternatives." Along with decreasing the credibility of such alternatives, the average test difficulty index might also be greatly decreased.

NA as the correct option. Obviously, if NA is never the correct option its credibility will suffer. Roid & Haladyna (1982) have suggested making NA the correct response in about 25% of the items which include it. Again, this percentage seems reasonable with 4-option items if we use a percentage which is the inverse of the number of options. The percentages of NA items used as "correct" varied widely: 41% to 100% (Dudycha & Carpenter, 1973; Tollefson & Tripp, 1986; Tollefson, 1987);10% to 15% (Hughes & Trimble, 1965; Mueller, 1975); and, about 25% (Wesman & Bennet, 1946). Forsyth & Spratt (1980) did not report what percentage of the time NA was "correct."

NA in mathematics items. Using NA as "correct" in mathematics items requiring calculations could easily invalidate the item, since one may select the "correct" NA option simply because one's miscalculation does not match any distractor. In four studies (Wesman & Bennet, 1946; Mueller, 1975; Forsyth & Spratt, 1980; Tollefson & Tripp, 1986) some mathematics items were used, but the extent of calculations required with items using NA as the correct response was not reported. The remaining three studies did not utilize mathematics items.

Selecting items for NA. The manner of selecting items for use with NA varied greatly in the cited studies. In two studies all items on the experimental test included NA (Wesman & Bennet, 1946; Forsyth & Spratt, 1980). The selection method was described only as "random" in one study (Tollefson & Tripp, 1986), as "subjective" in another (Hughes & Trimble, 1965), and was not indicated in two studies (Mueller, 1975; Tollefson, 1987). In the remaining study (Dudycha & Carpenter, 1973), it was stated simply that difficulty and discrimination indices were used to select items.

The present authors suggest that NA be used only in items whose resulting difficulty will likely be at a moderate and optimal level, since it is known suggested (Lord, 1953; Henrysson, 1971;

Ebel, 1979; Hopkins & Stanley, 1982) that maximum discrimination is possible only when difficulty is at such a moderate level. Given the general finding of the reviewed studies that NA decreases the difficulty index (makes item more difficult), it would seem reasonable to use NA only in items whose difficulty index is relatively high (an easy item), with the goal of moving item difficulty into the optimal range and possibly increasing discrimination. Ways of determining optimal difficulty are discussed in a later section.

Assessing NA's effects on difficulty and discrimination. In all but two of the seven studies, difficulty and discrimination indices were reported only on a test-wise basis in terms of means, rather than for each item. With such data one can only compare the mean changes in the difficulty and discrimination indices to assess the effect of using NA. Such a comparison would be valid if one could assume that item difficulty and discrimination are linearly related, that is, that a change in item difficulty in a given direction will always result in a particular directional change in item discrimination. However, since difficulty and discrimination are related in a non-linear fashion described by Lord (1953) or Hopkins & Stanley (1981), one could expect both increases and decreases in discrimination depending on whether the decrease in the difficulty index is toward or away from optimal difficulty. The increases and decreases in discrimination would then tend to cancel each other out when averaged together, and the effect of NA on discrimination would tend to be obscured. The present authors suggest that, due to the non-linear relationship between difficulty and discrimination, the effects of using NA must be examined on an individual "item-level" basis rather than by averaging these item parameters over a whole test. The two studies which reported data for each item were Wesman & Bennet (1946) and Tollefson (1987), and these data are re-analyzed in a later section with respect to optimal difficulty.

Proposed techniques

The "optimal difficulty approach". A curvilinear relationship between item difficulty and discrimination is suggested by the finding that maximum discrimination is possible only when difficulty is at a moderate level (Lord, 1953; Henrysson, 1971; Ebel, 1979; Hopkins & Stanley, 1981). For this paper "optimal difficulty" will be defined as a particular moderate value of the difficulty index, at which point the discrimination index can be a maximum. It is stressed that adjusting difficulty does not necessarily improve discrimination but merely encourages an item to reach its potential, which may or may not exceed the initial discrimination index (Thorndike, 1982). In this paper the practice of selecting an item for use with NA, if the item is among those with difficulty indices well above the optimal level is the central component of what will be referred to for brevity as the "optimal difficulty approach." The application of this approach will differ depending on whether one is a practitioner or a researcher. Both would follow the item-writing guidelines, but the researcher would also construct a list of the item parameters arranged in such a way that changes in the parameters can be easily described with respect to the optimal difficulty level. The use of such a list for an item-level analysis of NA is discussed in detail later.

Determining optimal difficulty. A simple method of determining optimal difficulty is to use difficulty and discrimination indices based on the upper and lower 27% scoring groups (Ebel,1979). In this approach item difficulty is the average proportion correct in the upper and lower groups and item discrimination is the difference between the upper and lower groups in terms of proportion correct. Using this approach optimal difficulty is always .50. This relationship has been depicted by Hopkins & Stanley (1981) and it can be shown mathematically that only when difficulty is .50 is it possible for discrimination to be 1.0.

Item difficulty may also be measured using the whole group of examinees' data and is referred to as $\mathbf{p}_i$, proportion correct for item **i**; likewise, the point biserial correlation, $\mathbf{r}_{pbis}$, is used as a discrimination index. As with the upper/lower groups measures, it is necessary to determine whether there is some optimal level of $\mathbf{p}_i$ associated with a maximum value of $\mathbf{r}_{pbis}$. Lord (1953) reported findings relevant to this matter in one of his early seminal papers on item response theory (IRT). Lord (1953) related $\mathbf{p}_i$ to θ_j, person j's ability estimated by an IRT model, and to the standard error of θ_j which is used as a measure of overall discrimination. As a measure of item discrimination he calculated the biserial correlation between one's raw score and θ_j. Lord (1953) found that when the item discrimination level is held constant for multiple-choice items, that the standard error of IRT θ_j is at a minimum when item difficulty level is "somewhat easier than halfway between the chance success level and 1.00." For 4-option and 5-option items these optimal levels of $\mathbf{p}_i$ were found to be .713 and .682, respectively.

Obviously, these IRT-based values cannot be generalized intact to a classical situation in which the proportion correct and point-biserial correlations are used. Two of the reasons are that IRT ability, θ_j, and one's raw test score are not linearly related, and, secondly, that in the present study the discrimination indices are not assumed to be equivalent for all items. Lord (1953) assumed that his items had equivalent item-test biserial correlations (used as discrimination indices). Hence, while the optimal difficulty level may be "somewhat easier than halfway between chance success and 1.0" we cannot necessarily use the specific optimal values found by Lord (1953). Hennryson (1971), in reference to Lord (1953), suggests that the average difficulty level of 5-option multiple-choice items should be "somewhere around .60." For the present study, it is proposed that optimal difficulty levels of .67 and .64 be used for 4-option and 5-option items respectively. These levels are about halfway between the chance success level and the level found by Lord (1953) to be associated with minimal standard errors of the IRT ability estimates.

Assessing the effects of NA. While "optimal difficulty" with respect to item discrimination is not a new idea, only two of the cited studies (Hughes & Trimble, 1965; Forsyth & Spratt, 1980) made any reference to such a relationship between the two parameters. Comparing average item difficulty to average item discrimination is analogous to that of calculating a Pearson correlation between two variables which have a curvilinear relationship. Likewise, just as a data plot is useful in determining the appropriateness of using a Pearson correlation, one may construct a list of the parameters for each item, arranged in order from highest to lowest conventional difficulty index values. In such a list it is clear where conventional item difficulty indices fall with respect to optimal difficulty. One can then easily compare the conventional and NA-format difficulty indices for each item, noting whether this change is toward or away from the optimal difficulty level. Also one can check the corresponding conventional and NA discrimination indices to see in which direction they have changed.

Since it is true that difficulty and discrimination are related in such a way that discrimination can achieve its maximum only when difficulty is at a moderate level, one would expect that (a) if the difficulty index becomes closer to the optimal difficulty level, then discrimination could be allowed to increase, and that (b) if the difficulty index becomes farther away from the optimal level, then discrimination could be forced to decrease. In the type of list of item parameters proposed one can easily tally the number of items in which either of these two changes has occurred. The greater the number of items with these changes, the more support for the notion that changes in item difficulty result, more often than not, in predictable changes in item discrimination, hence that lists of this type would be useful assessment tools.

That the item parameters relate in this predicted fashion, however, does not necessarily mean that our item-writing strategies have resulted in improved items. For example, for some items NA may cause a difficulty index to move farther from optimal and decrease discrimination. That is, while the literature suggests strongly that NA tends to decrease the difficulty index, in some cases when NA is used the difficulty index may increase away from optimal and reduce discrimination. While a decrease in discrimination would be predicted in these examples, it is not the desired psychometric outcome. Changes of this type may derive from unknown and/or uncontrolled factors affecting the item. Therefore, in the proposed lists, one should also keep a tally of the items which do improve, since that provides support that the item-writing rules used may be effective. The greater the number of items yielding this type of support, the more likely it is that following the specified item-writing rules helps to move difficulty towards optimal while increasing discrimination.

In this study the item parameters are listed in the proposed manner in Tables 1 through 5. In those tables is a heading entitled "Support" under which are two sub-headings entitled "NL" and "Rules." The title "NL" refers to the non-linear relationship between difficulty and discrimination just discussed in which discrimination can achieve its maximum only at moderate levels of difficulty. In this column "Y" (Yes) indicates that the changes in the item parameters would be predicted by that non-linear relationship; "N" (No) indicates the opposite. The sub-title "Rules" refers to the item-writing rules discussed earlier and "Y" (Yes) indicates that the changes in the item parameters are desirable or improved; "N" (No) indicates the opposite. The authors recognize that the use of these lists in this optimal difficulty approach is only a descriptive-level method, but at this point it seems much more appropriate than the averaging of item parameters used in previous studies to assess the effects of NA. In the re-analysis of data from Wesman & Bennet (1946) and Tollefson (1987) and in the pilot and main studies by the present authors, one would expect general support under the "Rel." column in these tables and increasing support under the "Rules" column as the certainty increases that the proposed item-writing rules have been followed in the use of NA.

Results

In this section, support is shown for using the optimal difficulty approach to investigate NA, first in terms of a re-analysis of the data from Wesman & Bennet (1946) and Tollefson (1987) and then from two studies by the present authors. In the studies by Wesman & Bennet (1946) and Tollefson (1987) the difficulty index was proportion correct in the whole group and the discrimination index was the item-test correlation or point-biserial correlation. Therefore, for those studies the optimal difficulty was based on Lord (1953) and Henrysson (1971). For the study by the present authors the difficulty and discrimination indices were calculated using both the whole group and upper/lower groups (Ebel, 1979) approaches.

The optimal difficulty approach was implemented by carrying out the following steps: (a) For each study, items were listed in order from greatest to least conventional difficulty index and it was noted where optimal difficulty fell within the ordered list, using an optimal difficulty level appropriate to the way item parameters were calculated and to the number of item options. The effect of NA was investigated by comparing the difficulty and discrimination indices of the NA form of each item with those parameters for the conventional form of the item. It was possible then to see how the discrimination index responded to the change of the difficulty index toward or away from optimal difficulty as a result of using NA; (b) largely because of the pilot study's item analysis it was possible to construct items for the main study in greater conformity with the item-writing suggestions discussed early in this paper. That is, item parameters were unknown in the pilot study, but parameter estimates were available for the main study. It was anticipated that following these suggestions would result in a greater percentage of the items showing increases in discrimination than was obtained in previous studies where the guidelines were not clearly or consistently followed.

Wesman & Bennet (1946)

In the study by Wesman & Bennet (1946), 591 applicants to nursing school were given a mathematics test and vocabulary test, each consisting of 20 5-choice items. About half of the applicants took tests in which the fifth option was "none of these" (treated as NA in this paper) and the remaining applicants took parallel tests with conventional fifth options. On the tests with NA, that option was the correct answer on five of the twenty items. For this study optimal difficulty was determined to be approximately .64 based on Henrysson (1971) and Lord (1953), as discussed earlier, since 5-choice items and whole-group indices of difficulty and discrimination were used. The twenty items were ranked and divided into items with conventional difficulty indices above or below the optimal level. Table 1 depicts the changes in difficulty and discrimination as a result of using NA as a fifth option on the Mathematics test. As can be seen in Table 1, under the sub-heading "NL" in 12 (60%) of the 20 items the item parameters changed in a manner predictable by a non-linear relationship between them.

Place Table 1 about here

As was pointed out earlier, the "support" indicated in Table 1 and others in this study does not necessarily mean that the item parameters were improved by using NA, but means instead that difficulty and discrimination changed as would be predicted in light of their non-linear relationship. For example, in Table 1 support is obtained from both Items 2 and 20, but discrimination improves (increases) only in Item 20 and decreases in Item 2. By contrast, the parameters for Item 1 in Table 1 do not change in the predicted directions, hence do not support the notion that discrimination can achieve its maximum only when difficulty is at a moderate level. Under the sub-heading "Rules" Item 12 was too close to optimal to use NA, leaving seven items above the optimal difficulty level to evaluate. Of those seven items, four (57%) had improved parameters (difficulty closer to optimal and discrimination increased) when NA was used.

In Tables 1, 2, and 3 under sub-heading "Rules" items are tallied as to whether NA appeared to improve the item parameters, despite the fact that the extent of use of item-writing rules by the authors is unknown. It is useful to compare the tallies under "Rules" for these studies with the main study by the present authors in which the proposed guidelines were carefully followed. If the proposed item-writing rules are valid, it is likely that this "Rules" tally for the authors' main study will be higher than that tally in studies where it is unlikely that all the guidelines were followed. Additionally, items whose original (non-NA) difficulty is below or very close to optimal difficulty are not evaluated under "Rules," since the proposed guidelines suggest using NA only with items whose original difficulty is well above the optimal level.

The results of the Vocabulary Test from Wesman & Bennet (1946) are presented in Table 2, and here similar results under "NL" were obtained in 11 (55%) of the 20 items. Under "Rules" only 3 (25%) of 12 items improved when NA was used. One may question the validity of the guideline suggesting that NA not be used as "correct" with mathematics items requiring calculations, since in this Wesman & Bennet study the "Rules" tally was much higher for the Mathematics test than for the Vocabulary test. However, this issue cannot be resolved since it was not reported as to which mathematics items required calculations.

Place Table 2 about here

Tollefson (1987)

In the study by Tollefson (1987) a test consisting of 73 4-option multiple-choice items was given to 81 students enrolled in a basic statistics course in education. No quantitative word problems were used as items, and 12 of the 73 items were used as the experimental items. The test was administered in three versions in which the fourth option in the 12 items was either (a) conventional, (b) NA as the correct answer, or (c) NA as a foil. Proportion correct was used as the measure of difficulty, and point-biserial correlations were used as the measure of discrimination.

An optimal difficulty level of .67 was used since these were 4-option items. The difficulty and discrimination indices for the 12 experimental items were then inspected in the manner described earlier. In Table 3 the data under "Support" has four columns of data instead of two as was found in Tables 1 and 2, because of Tollefson's use of each item as a foil or as the correct answer. Therefore, under "Foil" and "Correct" are the sub-headings of "NL" and "Rules." When considering NA used as a foil, and under the "NL" column, Items 1 and 3 were not included. The point-biserial correlation reported for Item 1 was .91, which is probably an error since the point biserial correlation is bounded by an absolute value of .80 when scores are normally distributed (Thorndike, 1982). Item 3 was not included, since there was no change reported for these item parameters. Hence, only 10 items were included in the analysis of support under "NL" using NA as a foil. It can be seen in Table 3 that 9 of the 10 items evidenced changes that would be predicted by the specified non-linear relationship between them. This tendency was much weaker when doing the same type of

---

Place Table 3 about here

---

comparison for NA as the correct answer in the corresponding sub-heading under "Correct." In that case only 4 of the 12 items supported the optimal difficulty notion.

Considering NA used as the correct option, Item 1 was not included for reasons already discussed. Also, Item 12 was not included because it was only .02 units higher than the optimal difficulty. Using NA with an item already so close to optimal could easily decrease the difficulty index to a level farther from optimal and thereby reduce discrimination. As a result, of the ten items remaining only two items were improved. In the corresponding analysis for NA as correct, only one item (9%) of the eleven were improved.

Overall, these findings from Tollefson's data follow the same pattern found in the other tables (including Tables 4 and 5)--more support for the non-linear relationship than for the item-writing rules. However, Tollefson's data seem more variable. This data must be interpreted cautiously since this tallying method used is, of course, purely descriptive and somewhat subjective. On the other hand, Tollefson breached three of the suggested item-writing rules by (a) substituting NA for the strongest distractor, (b) using NA in every item, and (c) using NA as "correct" with mathematics items requiring calculations. To the extent that these rules are valid, violations of them could be expected to result in low tallies under the "Rules" sub-heading. It is less clear how such violations might affect tallies under the "NL" sub-heading.

Pilot study by the present authors

Procedure and results. In the pilot study the authors utilized a 100-item, 4-option multiple-choice test taken by 300 undergraduate students as a final exam in Communication Fundamentals. To discourage cheating the test was given in two versions differing only in item order and differing slightly in the number of items using NA (six experimental items on one version and five on the other). If an experimental item used NA on one test version, that item appeared in a conventional format on the other test version, so that the NA and conventional formats could be compared in terms of item difficulty and discrimination. On each test version one of the experimental items used NA as the correct answer. In this pilot study no prior item statistics were available, and the items chosen for experimental use appeared to have a broad range of item difficulty as estimated by the instructor. The mean scores of the test versions were 64.18 and 66.01 and did not differ significantly. The two KR-20 reliabilites were .828 and .865 and also did not differ significantly ($F=1.27$, $df=145/153$, $p=.0722$) using a test by Feldt (1969). Difficulty and discrimination indices were based on the upper and lower 27% scoring groups (Ebel, 1979), hence a difficulty level of .50 was used as the optimal difficulty level for discrimination. The experimental items were ranked according to their conventional difficulty indices, as described earlier, and were examined with respect to optimal difficulty. In Table 4 it can be seen that 7 (64%) of the 11 experimental items behaved as would be predicted by the specified non-linear relationship. Under the "Rules" sub-heading 3 (38%) of the 8 items above optimal difficulty had improved item parameters when NA was used. This relatively low percentage should be viewed in light of the fact that no item analysis

---

Place Table 4 about here

---

was available for this pilot study, except for subjective estimates of difficulty by the instructor. Hence it was not known specifically how item difficulties ranked with respect to optimal difficulty, and the weakest distractors could not be adequately identified.

Main study by the present authors

Procedure. In the main study, the selection of items for use with NA was based on the item analysis of those items from the pilot study. The pilot study items were ranked according to conventional difficulty index, and 20 items from the highest 25% group were chosen as experimental items. Items were not selected if they asked the examinee to select the option which did "NOT" have some quality, because in conjuction with NA the examinee would have to negotiate a double negative (Osterlind, 1990). Two of the 20 items which were selected were later discarded by the instructor, because it was determined that the instructor had not covered the relevant material in class. Due to a clerical error one of the other experimental items was selected, although it had a relatively low prior difficulty index of .43. Therefore, there were 18 experimental items in this study. NA was substituted for the distractor which had been selected the fewest times by the group of examinees in the pilot study; this was also usually the weakest distractor in terms of point biserial correlations.

The test was administered to 337 undergraduates and consisted of 100 4-option items, administered in forms A and B, the latter differing from the former only in item order. The experimental items occurred in the same positions (assigned randomly) on each test; however, the NA option appeared only in the 18 experimental items in Version A, so that Forms A and B could be compared for the effect of NA on reliability.

Results. The results are presented in terms of how they bear on (a) the effect of NA on difficulty (using three different approaches--mean test scores, classical difficulty indices, and Rasch difficulties), (b) the effect of NA on discrimination with respect to optimal difficulty, (c) the size of the effect of NA on difficulty, (d) the attractiveness of NA when used as a distractor and correct answer, and (e) the effect of NA on test reliability.

As expected, the mean score (69.14) of the experimental version of the test was slightly lower ($t=2.14$, $p < .03$) than the mean score (71.22) of the conventional version. While the 18 experimental and 18 conventional items occupied the same positions on each test, item order differed on the versions in terms of the other 80 items. To help determine whether the difference in mean test scores was due to NA or to positional effects, the mean scores on the two test versions were calculated with the 18 items removed from each version. The resulting mean scores, 56.00 and 56.28, from the experimental and conventional forms respectively, were not significantly different, suggesting strongly (together with the earlier results) that the experimental test version was more difficult due to the presence of NA and not to positional effects.

The difficulty indices for the 18 experimental items were significantly lower than those for the 18 conventional items, when tested with a Wilcoxon matched-pairs signed ranks test ($Z=-3.1717$, $p=.0015$). This suggested, in agreement with the literature, that NA caused the experimental items to have lower difficulty indices.

Another test of the relative difficulty of the 18 experimental and conventional items was performed using difficulty measures from a Rasch analysis. While the classical difficulty indices, being proportions, require use of a non-parametric testing procedure, the difficulty measures produced in the IRT approach are considered to be interval level data and "sample-free" (Wright, 1960), which may permit their use with parametric procedures as discussed below.

While the n's (169 and 168) from this study did not permit using a two- or three-parameter IRT model, a one-parameter Rasch model (Wright, 1960) was used in order to obtain Rasch estimates of difficulty. Use of the Rasch model was justified, because it seemed reasonable to assume that ability was normally distributed, and because the "fit" of the model to the data was reasonably close as evidenced by Rasch item plots. If these assumptions are met, then the Rasch difficulties may be considered to be "sample free" (Wright, 1988). By "sample free" Wright (1988) asserts that "the difficulties of items can be compared even though they might come from quite different samples of persons."

The present authors suggest that this quality of being "sample free" may permit the use of Rasch difficulties in parametric procedures, such as a dependent t-test, to compare the difficulty levels of the 18 experimental and 18 conventional items. The reasoning is that, firstly, if the item is treated as the unit of analysis, the "sample-freeness" of the items may satisfy the independence assumption. Secondly, Wright (1988) states that IRT difficulty measures are at the interval level of measurement. Thirdly, since the 18 items in each group are matched (except that NA is used in one format) their difficulties are correlated. Accordingly, a Rasch analysis was conducted on the 98 items of the experimental and conventional versions of the test. Using a dependent t-test, the mean Rasch difficulty estimate (-.18) of the 18 experimental items was found to be significantly more difficult ($t=2.81, p=.012$) than the corresponding mean of the 18 conventional items (-.70). This finding parallels the non-parametric test of classical difficulty indices reported earlier.

To evaluate the "optimal difficulty" approach, the effect of NA on difficulty and discrimination was examined with respect to an optimal difficulty of .50, since the difficulty and discrimination indices were based on upper & lower 27% scoring groups (Ebel, 1979). In Table 5 under "NL" it can be seen that of the 17 instances in which the difficulty index changed as a result of using NA, 13 (76%) of those changes were as would be predicted by the specified non-linear relationship. Item 93 in Table 5 has a markedly lower difficulty index than the other items. As explained earlier, this was due to a clerical error in which that item was selected for the main study, despite having an initial difficulty index of only .47 in the pilot study.

Place Table 5 about here

The intent was to use only items with the highest difficulty indices above the optimal level. Nevertheless, this result supports the contention that NA used inappropriately may decrease item discrimination. Under "Rules" in Table 5 it can be seen that 10 (59%) of the 17 items above optimal difficulty improved when used with NA. Although not appreciably higher than the corresponding percentage of 57% found in Mathematics test from Wesman & Bennet (1946), this percentage of 59% is the highest percentage found in any of the studies in support of the item-writing rules. Clearly, more research is called for. As the conditions under which NA should be used are better defined, there should be increasing support of the type represented under the "Rules" subheadings in these tables.

Since the analyses in Tables 4 and 5 utilized indices based on upper and lower groups, a parallel analysis was done using proportion correct in the whole group as the difficulty index and using point-biserial correlations based on the whole group as the discrimination index. In that analysis the difficulty index remained the same for two items. In the remaining 16 items, the changes in difficulty and discrimination indices for 10 (about 62%) of the items supported the optimal difficulty approach. It was felt that in using the whole group of examinees, the mixture of high and low abilities in the middle 46% of the examinees probably made the optimal difficulty approach less sensitive than when used with only the upper and lower 27% scoring groups.

The effectiveness of NA as a distractor and correct option was evaluated by comparing the percentages of examinees selecting it with the percentages selecting the other three options. As a distractor NA was selected by 14.6% of the examinees, compared to options A (8.3%), B (5.9%), and C (6.5%). The percentages for positions A, B, and C were averaged across both versions of the test. On the conventional test, in which position D was a conventional option, option D was chosen only 3% of the time. It is suggested that this low percentage is due to the fact that in each experimental item, NA was substituted for the weakest distractor. This is strong support for NA as a replacement for weak distractors; however, it is unclear why NA was so much more attractive a distractor than the other three options. The usual method to control for position effects is to balance the key. This was done for the test as a whole and for the 18 experimental items in particular, such that each position contained the correct answer 24 or 25 times. Hence, the key was balanced.

As the correct answer, position D was selected by fewer examinees (66.9%) when NA was used than when it was in the conventional format (85.3%), results found also by Tollefson (1987) and Oosterhof & Coats (1984). The percentage of 85.3% is similar to the percentages selecting position A (88.7%) and C (82.8%). The percentage selecting position B was 64.8%; however, it is unclear why position B was associated with the lowest percentage, since the key was balanced. However, since this discussion concerns keyed options, the "attractiveness" reflects the general finding that NA decreases the difficulty index. In short, compared to conventional options, NA was less attractive as the correct response but more attractive as a distractor.

As was suggested earlier, since NA tends to decrease the difficulty index, and since optimal difficulty is associated with maximum discrimination and test reliability, NA should be used only with items having the highest difficulty indices above optimal difficulty so that the resulting difficulty level may be closer to and not much less than optimal. Therefore, an estimate of the size of the effect of NA on difficulty is necessary in order to estimate the result of using NA with any particular item whose initial difficulty is known. Also, since it has been reported that NA decreases difficulty more when NA is the correct option than when it is a distractor (Tollefson, 1987; Williamson & Hopkins, 1967), effect sizes using NA as a correct option and as a foil must be estimated separately.

To assess the magnitude of the effect of using NA as correct option and as a foil, the mean difficulty indices for the pilot and formal studies by the present authors, for Wesman & Bennet (1946), and Tollefson (1987) are presented in Table 6. The other studies reviewed did not report separate values for NA when used as correct and as a foil, hence their results could not be included. It can be seen in Table 6 that the effect of NA when correct is, in general, much larger than that when NA is a foil, which suggests that an overall effect size would be of little use. On average, the decrease in the difficulty index caused by a correct NA was 3.37 times the decrease caused by NA as a foil. For reasons discussed below, a better estimate may be obtained if the data from Tollefson (1987) and from the Mathematics test used by Wesman & Bennet (1946) are not used. When they are excluded, the effect of NA on difficulty when correct (.158), is about 2.29 times its effect as a distractor (.069).

In Tollefson (1987) the effect of NA on the difficulty index was the reverse of that of the remaining four studies, the mean difficulty index decreasing by .38 when using NA as a foil compared to a decrease of .23 when using it as the correct answer. This may derive in part from the fact that Tollefson used NA as correct on every item of the experimental test. If NA was obviously correct to many examinees the difficulty index would likely increase. Furthermore, Tollefson substituted NA for the strongest distractor in each item, which would make it even more likely for NA to be chosen. In short, since Tollefson's methodology was so much at variance with that used in the other studies, that it was not used to estimate the size of the effect of NA on difficulty.

Place Table 6 about here

The results from the Mathematics test used by Wesman & Bennet (1946) do result in a greater decrease in difficulty when NA is correct than when it is a foil. However, the size of these effects is much smaller than those in the other studies. Whether these relatively small effects are due to the use of Mathematics items or not is unknown, but because of possible problems with item validity discussed earlier when using Mathematics items, the results from this Mathematics test also was not used to estimate the effect size of NA on difficulty.

The results from the Vocabulary test of Wesman & Bennet (1946), however, are used to estimate the effect size of NA on difficulty, although they did not state explicitly that they substituted NA for the weakest option in their experimental items. It seems likely, however, that they did follow this practice since they stated, "If the option being removed is not very good, the none of these' option may prove of real value." It is likely, therefore, that their substitution was done either on a random basis or with the weakest distractor.

Using the data from Table 6 for the Vocabulary data from Wesman & Bennet (1946), and from the pilot and main studies by the present authors, it can be seen that the average percentage decrease in the difficulty index due to using NA was 20.4% when using NA as the correct answer and was 9.5% when using NA as a foil. Given that these percentages are only crude estimates of effect size at this point, one might use them to estimate whether using NA would move item difficulty sufficiently closer to the optimal level and hence be likely to increase discrimination. If the resulting difficulty index, however, is estimated to be farther from optimal difficulty, then one would be advised to not use NA with that item. The same type of procedure could be followed for estimating the effect when NA is used as a foil, using the appropriate estimate of the effect size.

Regarding the effect of NA on test reliability, the KR-20 reliabilities (.835 and .797, experimental and conventional, respectively) were compared using a test by Feldt (1969) and were not significantly different (F=1.23,df=168/167,p=.0908). It is noteworthy that the use of NA was, at least, not detrimental to test reliability, unlike results reported by Tollefson (1987). That Tollefson (1987) found a decrease in reliability associated with NA is not surprising given that that researcher replaced the strongest distractors with NA and also used NA on each item of the experimental test. Hence, attributing the decrease in reliability to NA does not seem justified. Moreover, when the KR-20 reliabilities reported by Tollefson (1987) were tested (Feldt, 1969) by the present authors, no significant differences were obtained. Others (Williamson & Hopkins, 1967; Forsyth & Spratt, 1980; Oosterhof & Coats, 1984) have reported mixed findings regarding the effects of NA on reliability. The results of those studies may be of questionable value, however, because in each study mathematics items requiring calculations were used and NA was used as the correct answer a certain percentage of the time. As suggested earlier in the paper, using NA with items requiring calculation may invalidate those items.

Summary and Discussion

With Haladyna & Downing (1989), the present authors agree that NA should probably not be used in a multiple-choice item if one can create enough good, conventional distractors. The present results are interpreted as supporting the use of NA when acceptable conventional distractors cannot be written, when previously unused distractors are of doubtful merit, or when weak distractors have been identified through item analysis. In such situations the present results support the judicious use of NA in light of optimal difficulty and by following other item-writing guidelines, with the result that NA may not only decrease the difficulty indices but may permit an increase in the discrimination indices of items in which NA is used. Test reliability did not suffer using NA under the conditions of this study.

It is suggested that these results may obtain under at least those conditions set up in the authors' main study, which were that (1) NA be substituted for the weakest distractor, that (2) the number of items using NA not exceed 20-25% of the total number (depending on the number of options), that (3) NA be used as the correct answer 20-25% of the time (depending on the number of options), that (4) NA not be used with mathematics items requiring calculations, and that (5) NA be used only on items for which there is reason to believe that the difficulty index is especially high and above the optimal difficulty level, that (6) NA be used only in items with clearly one answer, that (7) NA not be used with stems requiring a mental negating process since that, in conjunction with NA, would create an unnecessarily confusing double-negative situation, and that (8) NA be used once as the correct answer relatively early in the test to lend credibility to NA (see Williamson & Hopkins (1963), for example). It is the authors' opinion that probably guidelines 1, 5, and 7 have the most influence over the resulting difficulty level of an item, although the other guidelines may have more obvious relevance to the credibility of items using NA. Future research may investigate the relevance of these guidelines to the use of NA.

In studies of the effects of NA on difficulty and discrimination indices it is suggested that, along with following the above guidelines, the item parameters of items using NA be listed in order of original difficulty so that the effects of using NA may be evaluated with respect to the optimal difficulty level. The use of such an ordered list and the item-writing guidelines is referred to in this paper as the "optimal difficulty approach" to investigating the effects of NA on item difficulty and discrimination. In the present study consistent and fairly strong support was reflected in Tables 1 through 5 (under the sub-heading "NL") for the notion that discrimination tends to increase as difficulty approaches an optimal level and tends to decrease when difficulty departs from optimal. This finding supports the use of ordered lists as were used in these tables, rather than discussing results in terms of average difficulty and discrimination values, as was done in previous studies. Support was less firm and less consistent for the notion that item parameters may be improved (made somewhat harder and more discriminating) by following the item-writing guidelines. The clearest support in this regard came from Table 5 for the main study, the only study in which an item analysis was used in conjunction with the guidelines. Clearly, more research is needed to confirm or deny the usefulness of the present guidelines and to suggest other guidelines for using NA.

In the same vein, in both the pilot and main studies the authors found several items whose response to NA did not support the optimal difficulty approach. Some of the changes in parameters may have derived from random fluctuation and from comparing data from two independent though similar groups. Other factors might also have caused these items to respond to NA as they did. More research on the underlying mental processes elicited by NA in the examinee could be profitable. For example, it is suggested that the appropriate use of NA may raise the cognitive level of the item beyond the knowledge or recognition level. How does this enhancement take place? How do various mental processes, such as decision-making strategies for example, interact with factors such as number and type of options to affect difficulty and discrimination, and in which situations might NA be appropriate? In this regard, it seems likely that NA intensifies the effects of accompanying options and that conversely, as Wesman & Bennet (1946) suggested, the effect of NA depends mostly on the quality of the other options. It may also be instructive to compare the effects of substituting NA for the weakest distractor to the effects of merely making NA an additional option. Other relevant factors include the effects of classroom instruction, the content area, textbook characteristics, and student characteristics, all of which may be relevant to item construction and student response, especially with respect to the use of NA.

More research is needed to estimate the effect size of NA on the difficulty index. Since the size of the effect on difficulty of using NA is not clear at this time, it recommended that NA be used only with the items having the difficulty indices well above the optimal level, to avoid decreasing the difficulty indices so much that discrimination suffers.

Another consideration is whether one should use difficulty and discrimination indices based on using the whole test group or on upper/lower 27% scoring groups. The tendency is to recommend whole group measures since the upper/lower groups approach was developed primarily to alleviate the burden of calculations. While computer software has removed that burden, the present authors wish to raise the issue of whether the use of upper/lower 27% scoring groups, based as it is on "extreme groups," may be more sensitive than the whole-group methods for detecting items which need revision. While it has been shown that the correlation is very high between difficulty indices based on upper/lower groups and whole groups (proportion correct) (Michael, Haertzka & Perry, 1953) and between the discrimination index based on upper/lower groups and point-biserials based on whole groups (Beuchert & Mendoza, 1971), it is less clear which approach may best explore the relationship between difficulty and discrimination as affected by NA.

A further issue is how to determine the optimal difficulty level when using whole-group measures of difficulty and discrimination. The optimal values for proportion correct suggested by Lord (1953) were based on an IRT model which presumed items with equal discriminatory power. The point biserial correlations evaluated in the present study were not held constant, hence the optimal values for proportion correct were only subjective estimates based on Lord's work. Research using these classical item parameters might provide better estimates of optimal values for proportion correct. In addition to the use of classical item parameters, it is suggested that future research on NA with sufficiently large sample sizes may do well to utilize at least 2-parameter IRT models. One benefit would be that estimates of the effect size of NA on difficulty and discrimination might be assumed to be "sample free," hence permitting the effect size estimates to be more easily used with different types of samples.

In summary, while the issue of using NA in multiple-choice items is not of great theoretical significance, the authors suggest that this issue has practical significance, because of its widespread use and misuse.

References

Beuchert, A. K. & Mendoza, J. L. (1979). A monte carlo comparison of ten item discrimination indices. Journal of Educational Measurement, 16, 109-118.

Dudycha, A. L. & Carpenter, J. B. (1973). Effects of item format on item discrimination and difficulty. Journal of Applied Psychology, 58,116-121.

Ebel, R. L. (1979). Essentials of Educational Measurement (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, 149-150, 273.

Feldt, L. S. (1979). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson's coefficient twenty is the same for two tests. Psychometrika, 34, 363-373.

Forsyth, R. A. & Spratt, K. F. (1980). Measuring problem solving ability in mathematics with multiple-choice items: the effect of item format on selected item and test characteristics. Journal of Educational Measurement, 17, 31-43.

Haladyna, T. M. & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 2, 51-78.

Henrysson, S. (1971). Gathering, analyzing, and using data in test items. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education.

Hopkins, K. D. & Stanley, J. C. (1981). Educational and Psychological Measurement and Evaluation (Sixth Edition). Englewood Cliffs, New Jersey: Prentice-Hall, 272-273.

Hughes, H. H. & Trimble, W. E. (1965). The use of complex alternatives in multiple choice items. Educational and Psychology Measurement, 25, 117-127.

Lord, F. M. (1953). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. Psychometrika, 18, 57-77.

Michael, W. B, Haertzka, F. & Perry, N. C. (1953). Errors in estimates of item difficulty obtained from use of extreme groups on a criterion variable. Educational and psychological measurement, 13, 601-606.

Mehrens, W. A. & Lehman, I. J. (1984). Measurement and Evaluation in Education and Psychology (3rd ed.). New York: Holt, Rinehart, & Winston, 161-162.

Mueller, D. J. (1975). An assessment of the effectiveness of complex alternatives in multiple choice achievement test items. Educational Educational and Psychological Measurement, 35, 135-141.

Nitko, A. J. (1983). Educational Tests and Measurement, An Introduction. New York: Harcourt, Brace, Jovanovich, 206-207.

Osterlind, S. J. (1990). Constructing Test Items. Boston, MA: Kluwer Academic Publishers, 161-164.

Roid, G. H. & Haladyna, T. M. (1982). A technology for test-item writing. New York: Academic Press, 54.

Stanley, J. C. (1964). Measurement in Today's Schools (4th ed.). Englewood Cliffs, NJ: Prentice Hall, 238-239.

Thorndike, R. L. (1982). Applied Psychometrics. Houghton-Mifflin: Boston, p. 72.

Thorndike, R. L. & Hagen, E. (1969). Measurement and Evaluation in Psychology and Education (3rd ed.). New York: Wiley & Sons, 112-113.

Tollefson, N. (1987). A comparison of the item difficulty and item discrimination of multiple-choice items using the "none of the above" and one correct response options. Educational and Psychological Measurement, 47, 377-383.

Tollefson, N. & Tripp, A. (1986). The effect of an exclusion item format on item difficulty and item discrimination. (Report No. TM 011 080). Athens, Ohio: Ohio University, Alden Library. (ERIC Document Reproduction Service No. ED 292 810).

Weitzman, E. & McNamara, W. J. (1946). Apt use of the inept choice in multiple-choice testing. Journal of Educational Research, 39, 517-22.

Wesman, A. G. & Bennett, G. K. (1946). The use of 'none of these' as an option in test construction. The Journal of Educational Psychology, 46, 541-549.

Williamson, M. L. & Hopkins, K. D. (1967). The use of 'none of these' versus homogeneous alternatives on multiple-choice tests: experimental reliability and validity comparisons. Journal of Educational Measurements, 4(2), 53-58.

Table 1
Indices of Difficulty and Discrimination for a Mathematics Test

	Difficulty		Discrimination		Support	
Item	Conventional	NA	Conventional	NA	NL	Rules
* 8	.85	.74	.66	.64	N	N
* 3	.84	.74	.60	.61	Y	Y
* 2	.78	.65	.58	.62	Y	Y
20	.73	.80	.77	.69	Y	N
7	.71	.66	.43	.48	Y	Y
11	.70	.59	.53	.73	Y	Y
1	.69	.72	.28	.50	N	N
12	.66	.63	.65	.74	Y	?
4 ----	.64 ---	.59 ---	.68 ---	.73 --	Y -	**
19	.61	.63	.71	.64	Y	**
6	.48	.44	.66	.61	Y	**
5	.45	.49	.77	.81	Y	**
13	.41	.40	.78	.75	Y	**
9	.38	.34	.58	.68	N	**
*15	.29	.32	.52	.32	N	**
*16	.28	.37	.53	.27	N	**
14	.25	.30	.77	.75	N	**
10	.16	.32	.58	.51	N	**
18	.06	.08	.42	.28	N	**
17	.04	.06	.19	.26	Y	**

Note. Items split on optimal difficulty of .64. '*' -- correct response. Except for last two columns, data are from "The Use of 'None of These' as an Option in Test Construction" by A. G. Wesman & G. K. Bennett, 1946, The Journal of Educational Psychology, 37, p. 545-546. Data in public domain. '**' -- not assessed, below optimal difficulty. '?' -- too close to optimal.

Table 2
Indices of Difficulty and Discrimination for a Vocabulary Test

	Difficulty		Discrimination		Support	
Item	Conventional	NA	Conventional	NA	NL	Rules
1	.97	.99	.27	.17	Y	N
5	.92	.90	.47	.32	N	N
2	.90	.80	.58	.56	N	N
9	.87	.79	.11	.16	Y	Y
3	.80	.81	.36	.39	N	N
* 7	.80	.77	.50	.58	Y	Y
6	.74	.75	.65	.53	Y	N
*14	.74	.54	.58	.59	Y	Y
* 4	.73	.65	.64	.52	N	N
8	.73	.65	.48	.35	N	N
*11	.73	.48	.75	.59	N	N
13	.72	.66	.58	.45	N	N
17	.66	.51	.65	.65	N	?
12	.63	.56	.70	.64	Y	**
15	.59	.56	.63	.62	Y	**
10	.58	.45	.57	.55	Y	**
*16	.56	.47	.40	.14	Y	**
18	.47	.06	.30	.36	N	**
19	.46	.45	.68	.48	Y	**
20	.44	.37	.50	.45	Y	**

Note. Items split on optimal difficulty of .64. '*' -- correct response. Except for last two columns, data are from "The Use of 'None of These' as an Option in Test Construction" by A. G. Wesman & G. K. Bennett, 1946, The Journal of Educational Psychology, 37, p. 545-546. Data in public domain. '**' -- not assessed, below optimal difficulty. '?' -- too close to optimal.

Table 3
Indices of Difficulty and Discrimination for a Statistics Test

	Difficulty			Discrimination			Support			
		Foil	Correct		Foil	Correct	Foil		Correct	
Item	Conv.	NA	NA	Conv.	NA	NA	NL	Rules	NL	Rules
3	1.00	1.00	.89	.00	.00	.37	**	N	Y	Y
10	.92	.93	.67	.70	.24	.53	Y	N	N	N
9	.92	.89	.37	.24	.56	.55	Y	Y	Y	N
5	.88	.89	.33	.44	.56	.21	N	N	Y	N
7	.88	.85	.64	.33	.63	.28	Y	Y	N	N
2	.85	.93	.59	.55	.48	.41	Y	N	N	N
4	.81	.89	.52	.65	.23	.64	Y	N	N	N
11	.81	.86	.85	.69	.53	.38	Y	N	Y	N
1	.81	.61	.63	.91	.34	.62	**	**	N	N
6	.77	.39	.67	.74	.59	.50	Y	N	N	N
8	.73	.50	.52	.35	.31	.52	Y	N	N	N
12	.69	.43	.63	.71	.35	.38	Y	?	N	?

Note: All items exceeded optimal difficulty level of .67. Except for the last two columns data are from "A Comparison of the Item Difficulty and ITem DIscrimination of Multiple-choice Items Using the 'None of the Above' and One Correct Response Options" by Nona Tollefson, 1987, Educational and Psychological Measurement, 47, p. 380-381. Adapted by permission. '**' -- assessment was not appropriate. '?' -- conventional difficulty was too close to optimal for NA to be validly used.

Table 4

Effect of Using NA on Indices of Difficulty and Discrimination Based on Upper & Lower 27% Groups for Pilot Study of a Test of Communication Fundamentals

	Difficulty		Discrimination		Support	
Item	Conventional	NA	Conventional	NA	NL	Rules
*16	.93	.78	.10	.33	Y	Y
12	.87	.78	.21	.33	Y	Y
15	.77	.77	.46	.31	N	N
18	.69	.63	.43	.49	Y	Y
*21	.67	.45	.51	.38	N	N
19	.63	.67	.49	.29	Y	N
11	.62	.55	.26	.24	N	N
13	.58	.61	.28	-.07	Y	N
17	.46	.36	.56	.19	Y	**
20	.44	.47	.50	.54	Y	**
14	.19	.10	.10	.15	N	**

Note. Items split on optimal difficulty of .50. '*'- NA used as correct response. 'NL' -- support for non-linear relationship between parameters. 'Rules' -- support for item-writing rules. '**' -- not assessed, below optimal difficulty.

Table 5

Effect of Using NA on Indices of Difficulty and DiscriminationBased on Upper & Lower 27% Scoring Groups in a Test of Communication Fundamentals

	Difficulty		Discrimination		Support	
Item	Conventional	NA	Conventional	NA	NL	Rules
26	.98	.96	.04	.04	N	N
*43	.98	.98	.04	.04	**	N
36	.94	.93	.07	.09	Y	Y
*94	.94	.59	.11	.17	Y	Y
57	.93	.98	.04	.00	Y	N
21	.92	.88	.07	.11	Y	Y
47	.89	.90	.22	.15	Y	N
58	.88	.86	.16	.20	Y	Y
91	.87	.83	.27	.17	N	N
*92	.87	.68	.22	.46	Y	Y
8	.84	.72	.09	.30	Y	Y
67	.81	.32	.16	-.02	N	N
54	.79	.71	.29	.41	Y	Y
62	.79	.67	.42	.13	N	N
76	.77	.61	.16	.22	Y	Y
89	.72	.70	.16	.52	Y	Y
* 5	.63	.45	.20	.24	Y	Y
93	.24	.17	.22	.13	Y	**

Note. Items split on optimal difficulty of .50. '* '- NA used as correct response. 'NL' -- support for non-linear relationship between parameters. 'Rules' -- support for item-writing rules. '**' -- not assessed, below optimal difficulty.

Table 6

The Effect of "None of the Above" (NA) Used as Correct Option or Foil on Item Difficulty as Reported in Several Studies

Study	Number of Items	NA Usage	Conventional	NA	Change	Percent Change
Wesman & Bennet (1946)--Vocab	5	Correct	.712	.582	.130	-18.8
	15	Foil	.699	.621	.078	-11.2
Wesman & Bennet (1946)--Math	5	Correct	.606	.564	.042	-6.9
	15	Foil	.464	.470	-.006	+1.3
Rich & Johanson (Pilot)	2	Correct	.800	.615	.185	-23.1
	9	Foil	.583	.548	.035	-6.0
Rich & Johanson (Main study)	4	Correct	.855	.675	.180	-21.1
	14	Foil	.812	.731	.080	-9.9
Tollefson (1987)	12	Correct	.839	.609	.230	-27.4
	12	Foil	.839	.459	.380	-45.3

TM025616

AERA April 8-12, 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: AN ITEM-LEVEL ANALYSIS OF "NONE OF THE ABOVE"	
Author(s): CHARLES E. RICH & GEORGE A. JOHANSON	
Corporate Source: OHIO UNIVERSITY	Publication Date: APRIL 1990

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] **Check here** — Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

[ ] **or here** — Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: [signature]	Position: ASSOCIATE PROFESSOR
Printed Name: GEORGE JOHANSON	Organization: COLLEGE OF EDUCATION
Address: 201 McCRACKEN HALL, OHIO UNIVERSITY, ATHENS OH, 45701	Telephone Number: (614) 593-4487
	Date: APRIL 21, 1996

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 27, 1996

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1996/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://tikkun.ed.asu.edu/aera/). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

---

[1]If you are an AERA chair or discussant, please save this form for future use.