ED 307 281 TM 013 213

AUTHOR Engelhard, George, Jr.; And Others

TITLE Accuracy of Bias Review Judges in Identifying Differential Item Functioning on Teacher Certification Tests.

PUB DATE 11 Apr 89

NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Black Students; *Evaluators; Interrater Reliability; Item Analysis; Latent Trait Theory; *Licensing Examinations (Professions); Racial Bias; *Racial Differences; *Teacher Certification; *Test Bias; Testing Problems; Test Items; White Students

IDENTIFIERS *Accuracy; *Differential Item Performance; Review Panels

ABSTRACT

Wh^ther judges on bias review committees can identify test items that f .ction differently for black and white examinees was studied. Judges (n=42) on three bias review committees were asked tc examine a set of items and predict differential item functioning (DIF) without empirical data. Test items from teacher certification tests in the content fields of early childhood (n=11), administration and supervision (n=15), and middle childhood (n=16) were examined. Each commit^ee examined 40 items. Agreement between judgmental and empirical indices of DIF were determined. The results suggest tha' the agreement between the bias review judges and the empirical indices are generally not beyond what would be expected by chance, although each field had one to two judges who exhibited statistically significant agreement with the empirical indices of DIF. The data also indicate that the judges were unlikely to classify items as "favoring blacks." Suggestions for future research on the identification of biased items and the practical implications of this study are discussed. Five tables present the data. (Author/SLD)

ACCURACY OF BIAS REVIEW JUDGES IN IDENTIFYING DIFFERENTIAL ITEM FUNCTIONING ON TEACHER CERTIFICATION TESTS

George Engelhard, Jr.

Emory University

Linda Hansche and Kay Ellen Rutledge

Georgia Assessment Project

Georgia State University

Address: Professor George Engelhard, Jr.
Emory University
Division of Educational Studies
210 Fishburne Building
Atlanta, GA 30322

Running head: ACCURACY OF JUDGES

[Judges - Paper presented at the annual meeting of the American Educational Research Association, March 1989]

April 11, 1989

Abstract

The purpose of this study was to examine whether or not judges on bias review committees can identify test items which function differently for black and white examinees. Judges (n = 42) on three bias review committees were asked to examine a set of items, and predict differential item functioning without empirical data. Test items from teacher certification tests in the content fields of Early Childhood, Administration and Supervision, and Middle Childhood were examined here. Each committee examined 40 items, and agreement between judgmental and empirical indices of differential item functioning were determined. The results of this study suggest that the agreement between the bias review judges and the empirical indices are generally not beyond what would be expected by chance, although each field had 1 to 2 judges who exhibit statistically significant agreement with the empirical indices of differential item functioning. The data also indicate that the judges were unlikely to classify items as "favoring blacks". Suggestions for future research on identification of biased items and the practical implications of this study were discussed.

ACCURACY OF BIAS REVIEW JUDGES IN IDENTIFYING DIFFERENTIAL ITEM FUNCTIONING ON TEACHER CERTIFICATION TESTS

The analysis of test items for bias plays a critical role in the overall test development process. A variety of empirical methods have been used to identify items which function differently for certain groups of examinees (Berk, 1982; Cole & Moss, 1989), although the final decision to delete an item usually includes a consideration of both empirical data and the judgments by members of a bias review committee (Tittle, 1982). Typically, an empirical method, such as the Mantel-Haenszel Procedure (Holland and Thayer, 1988), is used to flag items which appear to perform differently for identifiable subgroups of examinees, and then a bias review committee makes its judgments on the basis of this empirical information in conjunction with other considerations.

This combination of empirical and judgmental information in the identification of biased items appears to work well for most testing programs where the number of examinees is large enough to obtain useful empirical estimates of group differences in item performance. However, there are a variety of testing programs which offer certification tests where the number of examinees is too small to justify the reasonable use of empirical methods to flag items which perform differently within relevant subgroups. For example, some content areas offered for teacher certification, such as foreign languages and special education, may have very few examinees. In these cases, the judges who are included on the bias review committees must make decisions regarding item bias without reliable empirical information. Previous research on the agreement between judgmental and empirical procedures has indic : that judges cannot accurately predict the items flagged by empirical indices of DIF (Pla'e, 1980; Rengel, 1986; Sandoval & Miille, 1980). Judgmental and empirical procedures tend to flag different items, and the unavailability of empirical data may be a significant problem in low-incidence fields.

In order to gain some insight into the potential problems which may be encountered in low-incidence certification fields, this study was designed to explore the extent to which judges can predict items which will perform differently for black and white examinees. Teacher certification tests in three content areas (Early Childhood, Administration and Supervision, and Middle Childhood) were selected because the sample sizes were adequate for obtaining empirical evidence of DIF which can be used to corroborate the judgmental predictions made by members of these three bias review committees. Although there are a variety of subgroups which can be examined for differential item functioning, this study focuses on differences between black and white examinees. Differences in item performance by race is an area which is of crucial concern because of the shortage of minority teachers, and the potential influence of certification tests in contributing to this shortage (Irvine, 1988).

This study differs in several important ways from previous research which has explored the agreement between empirical and judgmental methods for examining bias. One of the major differences is that the judges included in this study are the actual members of item bias review committees. These individuals are highly motivated professionals who were recommended by their colleagues for this judgmental task. Since the judges themselves are primarily teachers, it may be safe to speculate that they will be able to provide more accurate estimates of differential item performance than individuals who are not practitioners. Further, they received a 45 minute training session on item bias. Another difference is that much of the previous research was conducted using student data, while the current study uses items from teacher certification tests.

Two important methodological differences should also be noted. First, the judges in this study were asked to use three categories (favor blacks, no difference, favor whites) rather then simply biased versus nonbiased categories. Second, these judges were also asked to estimate the percentages of black and white examinees of comparable competence who would succeed on each item.

Purpose

The purpose of this study is to examine the agreement between the judgments of members of item bias review committees and an empirical assessment of differential item functioning on teacher certification tests. The specific research question addressed in this study is as follows: How well can judges predict which test items will perform differently for black and white examinees when they have no empirical information? Several exploratory analyses were also conducted to examine the relationship between the race of the judges (black/white) and selected aspects of the judgmental process.

Method

Subjects

Forty-two judges participated in this study. These judges were members of item bias review committees for teacher certification tests in the content fields of Early Childhood (n = 11), Administration and Supervision (n = 15) and Middle Childhood (n = 16). For the content field of Middle Childhood, there were originally 20 members on the committee, and 4 members were not included because of missing responses. A detailed description of the characteristics of the judges is presented in Table 1.

Insert Table 1 about here

Instruments

The test items which were examined for differential item functioning were drawn from teacher certification tests in the fields of Early Childhood, Administration and Supervision, and Middle Childhood. These test items are in multiple-choice format with 4 response categories per item. All of the items on each teacher certification test were classified on the basis of the Mantel-Haenszel Procedure into 3 categories (favor blacks, no difference, favor whites) using the chi-square statistic to determine statistical significance (alpha = .05), and the log of the MH summary estimate of the odds ratio to determine the direction of group differences. A table of random numbers was then used to select 40 items from each test with 10 items favoring blacks, 20 items with no evidence of group differences, and 10 items favoring whites. The item bias judges on the Administration and Supervision Committee examined 40 items. Due to errors in the printing of the items, judges in the content fields of Early Childhood and Middle Childhood examined 39 items. The deleted item in each case was in the no difference category. Each committee examined a different set of items drawn from the appropriate teacher certification test for their content field. There were no common items examined by members of different item bias review committees.

Procedures

The judges on each item bias review committee participated in a 45 minute training session. During this training session, the judges were presented with guidelines for identifying potentially biasing elements in test items. The judges were then asked to examine a set of 40 items without the benefit of any empirical information regarding differential item functioning, and to identify items which may perform differently for black and white educators. The specific questions were as follows: (1) Do you predict that this item will favor black or white educators of comparable competence? (favor blacks, no difference, favor whites), (2) What percentage of black and white educators of comparable competence will succeed on this item?, and (3) How confident are you in your prediction of differential item performance? (1 = low confidence to 6 = high confidence). The judges were then asked to comment on why they predicted that an item may bias the performance of either group.

The responses to question (1) were used to define a categorical index of DIF called the Judged Category (JCAT) Index with categories coded as follows: -1 = favor blacks, 0 = no difference, 1 = favor whites. The responses to question (2) were used to define a quantitative index of DIF which corresponds to the log odds ratio (Fleiss, 1981) called the Judged Log Odds Ratio (JLOR) Index. The JLOR Index was calculated as follows: $\ln[Pw/(1-Pw)] - \ln[Pb/(1-Pb)]$, where Pw is proportion of white examinees judged to succeed on the item, and Pb is the corresponding proportion for black examinees. Two comparable indices were obtained from the Mantel Haenszel Procedure. An Empirical Category (ECAT) Index was obtained using the MH chi-square statistic and the log of the MH odds ratio as described earlier to obtain three categories (favor blacks, no difference, favor whites). The Empirical Log Odds Ratio (ELOR) Index is simply the log of the weighted estimate of the odds ratio for whites obtained from the MH Procedure.

The percent agreement between the judgmental index of categorical DIF (JCAT), and the empirical index of categorical DIF (ECAT) were computed. Kappa statistics were also calculated to provide an index which is corrected for chance agreement (Cohen, 1960; Fleiss, 1981). When there is complete agreement between the judgmental and empirical indices, kappa = 1; if the agreement is greater than chance, kappa > 0 and if the observed agreement is less than or equal to chance, kappa <= 0. These statistics were also calculated separately for each category as recommended by

Fleiss (1981); for example, the agreement index for the no difference category was computed by combining the favor blacks and favor whites categories versus the no difference category. The critical ratio statistic proposed by Fleiss (1981) was used to test the statistical significance of the individual kappa statistics (alpha = .05) for each judge.

Pearson correlations were computed and used to examine the agreement between the judgmental estimates based on the judged log odds ratio (JLOR Index) and the empirical estimates obtained from the MH Procedure (ELOR Index).

Results

The distribution of the percent agreement between the classification of the items by the judges (JCAT) and the MH Procedure (ECAT) are presented in Table 2 for each content field.

Insert Table 2 about here

The medians range from 46.2 to 50.0 percent agreement. The summary information for the kappa statistics are presented in Table 3.

Insert Table 3 about here

The medians range from .02 to .09 with the Middle Childhood judges displaying the greatest average agreement with the empirical categorical index (ECAT). One judge in Early Childhood exhibited a statistically significant level of agreement after correcting for chance, kappa = .27. Two judges in Administration and Supervision exhibited significant levels of agreement with kappas of .18 and .14, while there was 1 judge in Middle Childhood who exhibited a significant level of agreement with a kappa of .18.

An examination of category usage indicates that the judges were unlikely to classify any of the items as "favoring blacks". The percent of responses in each of the three categories (favor blacks, no difference, favor whites) respectively were 1.4, 83.7 and 14.9 for Early Childhood; .8, 86.7 and 12.5 for Administration and Supervision; and finally, 1.3, 85.2 and 13.5 for Middle Childhood.

The percent agreement and kappa statistics for each category are presented in Table 4. The percent agreement between the judgmental and empirical classification of these items tends to be fairly high for the favor blacks category with median values ranging from 74.3 to 75.0. The kappa statistics indicate that this high agreement may be misleading, and due to the infrequent usage of the favor blacks category; median values of the kappa

Insert Table 4 about here

statistics are equal to zero across the three fields. The no difference category exhibits the next highest percent of agreement with medians ranging from 69.2 to 72.9 with the median of the kappa statistics ranging from .04 to .06. The final category of favor whites has the lowest percent agreements with medians ranging from 48.7 to 52.6 across fields. The median kappa statistics show less agreement across fields with average values ranging from -.00 for Early Childhood through .05 for Administration and Supervision to .07 for Middle Childhood. As might be expected, due to the infrequent usage of the favor blacks category by these judges, the no difference versus biased items (favor blacks and favor whites categories combined) exhibit the best agreement across fields.

In addition to the agreement between the two categorical indices of DIF (JCAT and ECAT), the Pearson correlations between the two quantitative indices of DIF based on the log odds ratios (JLOR and ELOR) were computed and are presented in Table 5.

Insert Table 5 about here

The median Pearson correlations ranged from .00 to .11. The distributions suggest that there are significant individual differences in judge accuracy with one judge in Early Childhood being able to predict DIF fairly accurately, r = .52, while one of the judges in Administration and Supervision had a substantial negative correlation, _r_ = -.36. These indices may also reflect judge engagement in the task. For example, the within judge agreement between the JCAT and JLOR indices varied by judge, and for the Administration and Supervision judge this correlation was quite low, _r_ = .28. The correlations obtained with the quantitative indices of agreement between judgmental and empirical DIF support the findings obtained with the categorical indices of agreement.

Exploratory Analyses

Although the sample sizes are small, several exploratory analyses were conducted to examine the relationship between the race of the judges (black/white) within each committee and selected aspects of the judgmental task. The kappa statistics, Pearson correlations, and percent of items judged to be biased (favor whites and favor blacks categories combined) for each judge were transformed to linear scales before the _t_ tests were conducted.

In Early Childhood, the black judges did not exhibit significantly higher average kappa statistics (_M_ = .08) than the white judges (_M_ = -.03), _t_ (8) = 2.18, _ns_. When the Pearson correlations are used to measure agreement, the average difference between the black (_M_ = .12) and white (_M_ = .07) judges was also not statistically significant, t (8) = .36, ns. Race does appear to be related to the percent of items classified as biased with black judges (M = 10.3) indicating fewer items than white judges (M = 19.0), t (8) = 2.30, p < .05. The mean reported level of confidence for th: black judges (M = 4.6) was not significantly different from the white judges (M = 4.2), t (8) = 1.10, ns.

No significant differences were found between black and white judges who were members of the Administration and Supervision Committee. The black judges did not exhibit significantly higher average kappa statistics (M = .06) than the white judges (M = -.00), t (12) = 1.91, ns. The average difference between the Pearson correlations for the black (M = -.00) and white judges (M = -.02) was also not significant, t (12) = .13, ns. The average percent of items classified as biased by black judges (M = 17.2) is not statistically different from the average for white judges (M = 12.5), t (12) = -.70, ns. The average degree of confidence was also similar for the black (M = 4.9) and white (M = 4.4) judges, t (12) = .96, ns.

There were also no significant differences related to the race of the judges in Middle Childhood. The black judges did not exhibit significantly higher average kappa statistics (M = .08) than the white judges (M = .09), t (14) = -.94, ns. The mean differences between the Pearson correlations for the black (M =

.10) and white judges (M = .07) were also not significant, t (14) = .46, ns. Race does not appear to be related to the average percent of items classified as biased by black judges (M = 18.7) as compared to the white judges (M = 9.6), t (14) = -1.37, ns. The average degree of confidence was also not significantly different for the black (M = 4.2) as compared to the white judges (M = 3.5), t (14) = 1.46, ns.

In summary, the results of the exploratory analyses suggest that the differences between the black and white judges are minimal with the exception of the percent of items classified as biased by the judges in Early Childhood where black judges classified fewer items as biased than white judges.

Discussion

The results of this study suggest that judges cannot predict which test items will perform differently for black and white examinees when they have no empirical information to guide their judgments. In each content field, there were only one to two judges with better than chance agreement, although the strength of agreement was still slight. According to the descriptions proposed by Landis and Koch (1977) for interpreting kappa statistics, one judge in Early Childhood exhibited fair agreement, while the other judges with significant kappa statistics reflect slight agreement. The quantitative indices of agreement between judgmental and empirical DIF also indicate that these judges cannot predict differential item functioning very well. The exploratory analyses suggest that the differences between the black and white judges on selected aspects of the judgmental process are minimal.

There are a number of strengths and limitations associated with this study that must be considered before interpreting the results. The major strength of this study is that judges are actual members of bias review committees. These judges were highly motivated, participated in a 45 minute training session, and were carefully selected because of their sensitivity to bias issues. A second strength is that the judges were given the opportunity to estimate the percent of comparable black and white examinees who would succeed on each item, rather than being asked to simply classify an item as biased or not biased. One of the limitations of this study is that the strength of the agreement between the judges and empirical indices of DIF may be attenuated by several factors. This low agreement is related to the infrequent use of the "favor blacks" category by these judges. Agreement may also be underestimated because these items have already been extensively screened for bias at earlier stages of the test development process, and many of the obvious sources of bias that might otherwise be observable to the judges have already been eliminated. Another factor which may lower the agreement is the reliability of the each of the indices. Further research is needed on the reliability of judges on bias review committees, as well as the reliability of the empirical methods used to identify DIF. Finally, the exploratory analyses regarding black and white differences are based on small sample sizes, and the findings need to be confirmed with additional research.

With these strengths and limitations in mind, the results of this study indicate that the agreement between the judgmental and empirical indices of DIF are very low and usually not better than what would be expected by chance. Although the results of this study confirm earlier findings regarding the accuracy of judges, the question still remains of who is a "good" judge. This question cannot be answered simply in terms of agreement between judgmental and empirical indices of DIF. One plausible interpretation for the low agreement found in this study is that the judgmental and empirical indices measure different aspects of item bias. As pointed out by Shepard (1982), item bias can be conceptualized as invalidity which distorts the meaning of the test results for some groups. Complementary evidence from both judges and empirical methods can contribute to our understanding of what the test scores mean and whether or not this meaning is confounded by irrelevant factors related to bias.

This study was motivated by a concern with problems related to the identification of item bias in content fields with small numbers of examinees. The data reported here suggest that the judges and empirical methods are providing different information regarding differential item functioning. Considering the nature of the judgmental task, especially when little or no reliable empirical data is available, further research is needed on how to identify and train judges who can assist test developers in identifying items which may bias the performance of certain examinees. Several avenues for future research seem promising. One approach would be to develop a set of items with known bias structure, and use this instrument to examine the ability of the judges to identify different types of item bias. This set of items could also be used to evaluate item bias training sessions, as well as provide a means for eliminating some judges.

Individual differences among judges may also be an important factor related to the quality of judgments. All judges may not be equally sensitive to item bias. The judgmental task is very demanding, and the judges are asked to represent the interests of their social category (race, gender) in a high stakes situation. These circumstances may be stressful for some judges. Anxiety about the performance of the task is probably increased when the judges do not have empirical data. Experience on item bias committees and training sessions may alleviate some of these problems. Further research may indicate that a core group of 5-10 "gcod" judges be included on each bias review committee in addition to the content area experts.

In summary, this study has perhaps raised more questions than it has answered concerning the role of judges in the identification of item bias. The data suggest that it is probably unreasonable to expect judges to flag the same items which are identified by empirical procedures. Further, even though the judges exhibit low agreement with an empirical procedure, it does not follow immediately that the quality of the judgments are low. Additional research is needed on defining the characteristics of a "good" judge for bias review committees regardless of whether or not reliable empirical information on differential item functioning is available.

References

Berk, R. A. (1982). (Ed.). Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cole N. S. & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), Educational Measurement. Third Edition. New York: Macmillan Publishing Company.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. Second Edition. New York: John Wiley & Sons.

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity, (129-145). Hillsdale, NJ: L. Erlbaum Associates, Publishers.

Irvine, J. J. (1988). An analysis of the problem of disappearing black educators. The Elementary School Journal, 88, 503-513.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.

Plake, B. S. (1980). A comparison of a statistical and subjective procedure to ascertain item validity: One step in the test validation process. Educational and Psychological Measurement, 40, 397-404.

Rengel, E. (1986). Agreement between statistical and judgmental item bias methods. Paper presented at the annual meeting of the American Psychological Association, Washington, DC. (ERIC Document Reproduction No. ED 289 890)

Sandoval, J. & Miille, M. P. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. Journal of Consulting and Clinical Psychology, 48, 249-253.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 9-30). Baltimore: Johns Hopkins University Press.

Tittle, C. (1982). Use of judgmental methods in item bias studies. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 31-63). Baltimore: Johns Hopkins University Press.

Table 1

Description of the Judges by Content Field

	Early Childhood (n = 11)	Administration/ Supervision (n = 15)	Middle Childhood (n = 16)
Ethnicity			
Black	5	8	11
White	5	6	5
Am. Indian/ Alaskan Native	1	0	0
Asian/Pacific Is.	0	1	0
Gender			
Male	2	6	2
Female	9	9	14
Age			
22-35	2	0	4
36-55	9	13	12
Over 55	0	2	0
Current Assignment			
Teacher	10	2	14
Administrator	0	11	1
Instr. Supervisor	1	1	0
Other	0	1	1
Committee Experience			
Yes	10	10	11
No	1	5	5

Table 2

Distributions of Percent Agreement

			Administration/			
	Early Childhood		Supervision		Middle Childhood	
Stem	Leaf	Freq.	Leaf	Freq.	Leaf	Freq.
6						
6						
5	9	1	55	2		
5	1	1	0000022	7	000111134	9
4	66999	5	5577	4	689999	6
4	114	3	2	1	1	1
3	8	1	5	1		
3						
2						
2						
Median	=	46.2		50.0		50.0
SIQR	=	3.8		3.8		1.3
Mean	=	46.6		48.5		49.5
SD	=	5.7		5.2		2.9
N	=	11		15		16

Note. SIQR is the semi-interquartile range.

Table 3

Distributions of Kappa Statistics

	Early Childhood		Administration/ Supervision		Middle Childhood	
Stem	Leaf	Freq.	Leaf	Freq.	Leaf	Freq.
.3						
.3						
.2	7	1				
.2						
.1			8	1	58	2
.1			34	2	11222	5
.0	799	3	7779	4	689	3
.0	23	2	14	2	0001	4
-.0	1	1	000	3	00	2
-.0	5566	4	6	1		
-.1			13	2		
-.1						
Median	=	.02		.04		.09
SIQR	=	.07		.05		.06
Mean	=	.03		.03		.07
SD	=	.10		.09		.06
N	=	11		15		16

Table 4

Percent Agreement and Kappa Statistics by Category and Content Field

Field		Favor Blacks		No Difference		Favor Whites	
		Agree	Kappa	Agree	Kappa	Agree	Kappa
Early Childhood (N = 11)							
Median	=	74.3	.00	69.2	.04	48.7	-.00
SIQR	=	1.3	.02	3.8	.17	5.1	.10
Mean	=	73.4	-.01	70.6	.10	49.2	.00
SD	=	1.7	.06	7.7	.20	5.7	.11
Administration and Supervision (N = 15)							
Median	=	75.0	.00	72.5	.04	52.5	.05
SIQR	=	.0	.00	5.0	.07	3.8	.08
Mean	=	74.5	-.00	69.5	.02	53.0	.06
SD	=	2.2	.05	6.2	.12	6.8	.14
Middle Childhood (N = 16)							
Median	=	74.3	.00	72.9	.06	52.6	.07
SIQR	=	1.2	.05	2.3	.08	4.4	.08
Mean	=	75.0	.03	70.6	.09	53.4	.08
SD	=	1.7	.07	6.6	.10	5.6	.11

Table 5

Distributions of Pearson Correlations

Stem	Early Childhood		Administration/ Supervision		Middle Childhood	
	Leaf	Freq.	Leaf	Freq.	Leaf	Freq.
.6						
.5	2	1				
.4						
.3	0	1	0	1		
.2	5	1	6	1	178	3
.1	0056	4	014	3	46777	5
.0	47	2	000	3	3589	4
-.0			679	3	00	2
-.1	8	1	47	2	46	2
-.2			3	1		
-.3	0	1	6	1		
-.4						
Median	=	.10		.00		.11
SIQR	=	.10		.13		.08
Mean	=	.11		-.01		.10
SD	=	.22		.18		.13
N	=	11		15		16

Note. Pearson correlations between quantitative indices of differential item functioning obtained from judges and Mantel-Haenszel procedure.