IMPLEMENTING THE CRITERION REFERENCED USAF APPRENTICE KNOWLEDGE TEST PROGRAM

1Lt Christopher M. Antons
USAF Occupational Mesurement Center

An Air Force Apprentice Knowledge Test (AKT) is designed to measure specialty knowledge at the three-skill or apprentice level of a specific Air Force enlisted specialty. The Occupational Test Development Branch of the USAF Occupational Measurement Center (USAFOMC) is responsible for developing and maintaining the AKTs. AKTs are used in conjunction with other factors to select airmen for bypass of technical training and direct entry into a specific career field at the apprentice level.

Prior to 1976, AKTs were 65-item multiple choice tests with passing scores set annually at the thirtieth percentile of the score distribution for all examinees who had previously taken a specific AKT. Inherent to the method, passing scores fluctuated, sometimes dramatically, depending upon the examinee population for a given year. AKT use in some specialties was very low, thereby severely limiting the reliability of the passing score. Conversely, for high usage AKTs, a change of only one point for the calculated passing score on this relatively short test meant a large difference in the total number of airmen passing or failing. For the few 65-item AKTs still in existence, the passing scores range from 19 to 26 raw score points. The most severe limitation of this method was the fact that the passing scores were established relative to the examinee population without reference to job incumbents or expected performance.

To improve the AKTs, USAFOMC initiated a series of studies. In the first study, AKT scores were compared for three groups: beginning trainees and graduates of a technical training course for general vehicle maintenance, and airmen already selected for bypass in that specialty, (Vaughan, 1976a). Mean scores for both the beginning trainees and bypass group were significantly lower than the mean score for graduates. Differences in scores of beginning trainees and graduates showed the test was able to discriminate among levels of knowledge for a specialty. Differences in scores of the bypass group and graduates demonstrated specialty knowledge differences between a group seeking apprentice skill level and a group just completing formal technical training. In comparison, the score at the tenth percentile of graduates was the same as the score at the seventy-fifth percentile of the bypass group. Using the score just above the tenth percentile as a passing score, some airmen previously selected as bypass specialists would not be qualified.

A second study replicated the first study on an additional five Air Force specialties and found similar results (Vaughan, 1976b). Based on the results of these studies, the USAFOMC implemented a criterion referenced testing program for AKTs using technical training graduates as the criterion group and the tenth percentile of that group as the passing score. The rationale given for originally setting the passing score above the tenth percentile was that extremely low scores are likely to contain considerable error (Lord and Novick, 1978). Conversely, a higher passing score was decided against since it might prevent acceptable performers from being selected to bypass training. Performance of technical school graduates and selected bypass specialists from one of the five specialties in the previous study were compared (Vaughan, 1978). Performance of the bypass specialists was shown to be equal to or slightly better than the technical school graduates. This evidence supported the decision not to set the passing score any higher than just above the tenth percentile.

In 1978, the USAFOMC began converting all AKTs to 100 items and criterion referencing those tests with a high usage (greater than 25 administrations per year), and a large enough criterion group of technical school graduates. All AKTs were expanded to

100 items to increase their reliability. The criterion referencing anchored the performance of bypass specialist candidates on the AKT to a known level of performance of technical school graduates. This allowed us to assume that successful bypass candidates had at least as much knowledge as the lower ten percent of technical school graduates for a given specialty.

Two main problems were encountered. First, we assumed that a few members of the examinee group would lack motivation for testing since they had just graduated, were preparing to depart for duty assignments, and were aware that the test had no impact on their own training. The USAFOMC explained the significance of this testing to training personnel and, in turn, the graduating trainees. This helped dispel the motivation problem. The second problem involved subject-matter experts (senior noncommissioned officers brought to the USAFOMC from working units in each specialty to provide input on content of the tests). They wanted to increase the difficulty of the tests to insure that bypass specialists would be knowledgeable. Test developers at the USAFOMC explained that increasing the difficulty of the test would also decrease the average score of the criterion group. With a lower mean criterion score, the passing score would be set lower. If set low enough, some examinees might achieve a passing score by chance alone.

Current Status

As of September 1982, of 270 specialties the following number of AKTs are available.

TYPE	NUMBER OF SPECIALTIES	AVERAGE USAGE
Criterion referenced	87	86
Noncriterion referenced	45	25
No test	138	--

The AKT program now includes both criterion and noncriterion referenced tests. All AKTs are criterion referenced unless annual usage is too low to justify the criterion referencing. For many specialties, no AKT is constructed because training is mandatory or other reasons specific to the individual specialties.

The following table provides information on usage of AKTs. Airmen take the examinations to bypass technical training when first entering the service (bypass) or when changing from one specialty to another (retraining), or to demonstrate apprentice level competency after a period of on-the-job training (upgrade).

AKT UTILIZATION
(CY 81)

USE	TOTAL TESTED	PERCENT PASS
Bypass	3517	66%
Retraining	1771	82%
Upgrade	1951	84%
Total	7299	75%

(Jan-Jun 82)

USE	TOTAL TESTED	PERCENT PASS
Bypass	934	59%
Retraining	1334	80%
Upgrade	962	83%

The following table provides information on the passing scores established for both the criterion and noncriterion referenced AKTs.

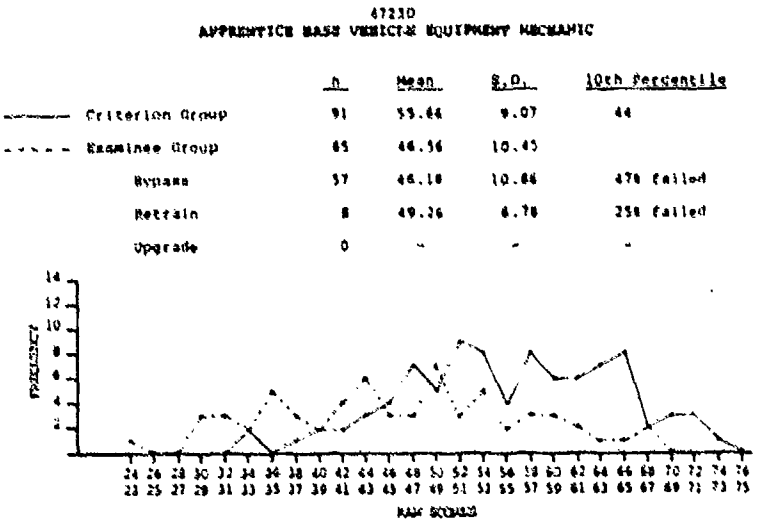
PASSING SCORE DISTRIBUTIONS

	N	Mean	SD	Range	%Passing
Criterion referenced	81	42.96	8.37	26-60	78%
Noncriterion referenced	29	42.28	6.20	30-56	77%
(65 item)Noncriterion referenced	6	24.50	2.74	19-26	85%

The average passing scores are nearly the same for criterion and noncriterion referenced tests. The difference, then, is not in placement of the passing score, but in the criterion that determines that score and the distribution of scores for that criterion group. As will be shown later, score distributions for the criterion groups are much less varied than for the examinee groups. Also, for % of examinees achieving passing scores, criterion and noncriterion referenced tests are nearly the same. According to the criteria for setting the passing score on noncriterion referenced tests, only 70% of examinees should pass. However, as stated earlier, the passing scores can fluctuate from year to year according to the population of examinees and the number of examinees passing depends upon the distribution of scores for one group compared to all past examinees. For the 65-item noncriterion referenced tests, there is more opportunity for fluctuation in scores from year to year and for examinees to achieve passing scores by chance.

Some Specific Criterion Referenced AKTs

Six criterion referenced AKTs were selected for analysis of both the criterion group and examinee group scores. Analyzing the AKTs individually, two were from specialties previously studied by Vaughan (1976a, 1978), two were selected for having extremely low passing rates and two were selected for having extremely high passing rates. Air Force Specialty Code (AFSC) 47230, Apprentice Base Vehicle Equipment Mechanic is similar to the general mechanic specialty examined by Vaughan (1976a). The passing score of 45 on this test is close to the average of 43 for all criterion referenced AKTs. In the 1976 study, the tenth percentile of the criterion group was the 75th percentile of the bypass group. In this case, the tenth percentile of the criterion group is the 47th percentile of the bypass group and the test is much more selective for the bypass group than the retraining group. For the purposes of the AKT, these characteristics are desireable.

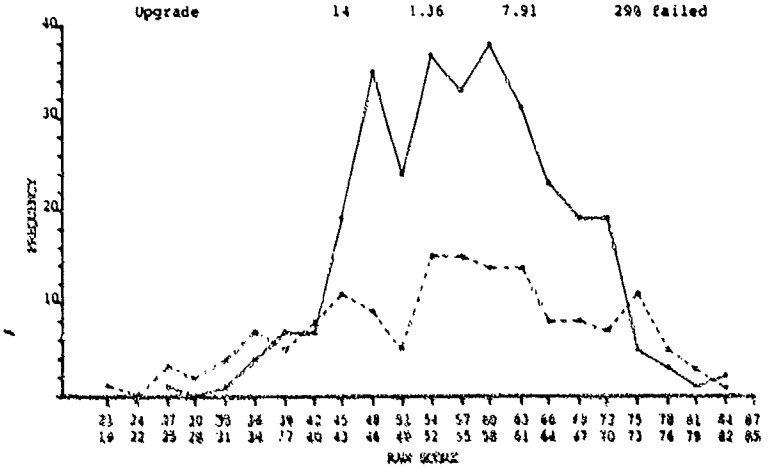


AFSC 90230, Apprentice Medical Service Specialist, was used in the performance measurement study (Vaughan, 1978) and criterion referencing study (Vaughan, 1976b). The passing score of 46 is also near the average for all criterion referenced AKTs. 33%

of the bypass group were below passing on this test compared to 58% in the 1976 study. Though means of examinee and criterion groups were nearly the same, the examinee group had the greater variance. The variance was not due to subgroups, since bypass and retrain groups both had large variance with their standard deviations twice the difference of their means. The upgrade group had less variance but was a smaller group and had a mean similar to the bypass group. What was notable was the large percent passing in the bypass group, indicating that, for this career field, civilian experience may provide adequate background. Considering the variance of the examinee groups, the cutoff scores were able to discriminate among examinees despite the similarity of examinee and criterion mean scores.

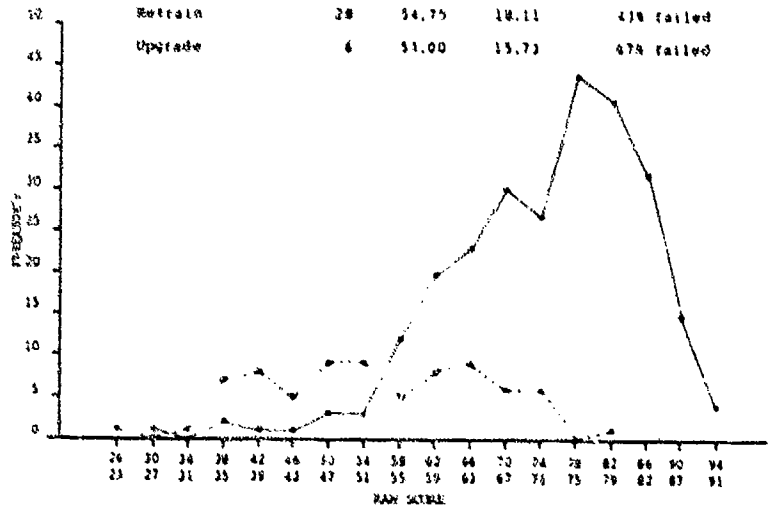
90230
APPRENTICE MEDICAL SERVICE SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	309	56.34	9.72	45
- - - - - Examinee Group	170	54.41	14.69	
Bypass	79	51.28	14.63	33% failed
Retrain	77	58.18	14.94	20% failed
Upgrade	14	1.36	7.91	29% failed



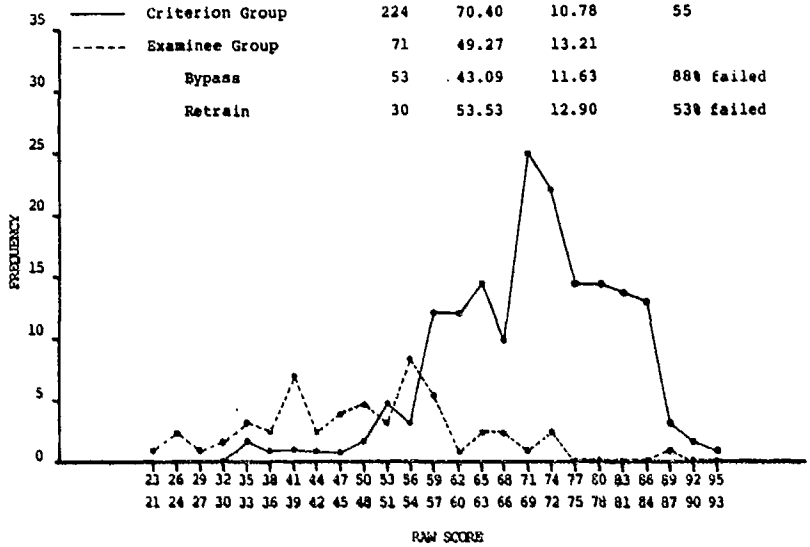
55235
APPRENTICE PLUMBING SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	259	73.05	10.65	59
- - - - - Examinee Group	76	52.33	14.27	
Bypass	42	50.62	11.22	55% failed
Retrain	28	54.75	18.11	43% failed
Upgrade	6	51.00	15.73	67% failed



55330
APPRENTICE ENGINEERING ASSISTANT SPECIALIST

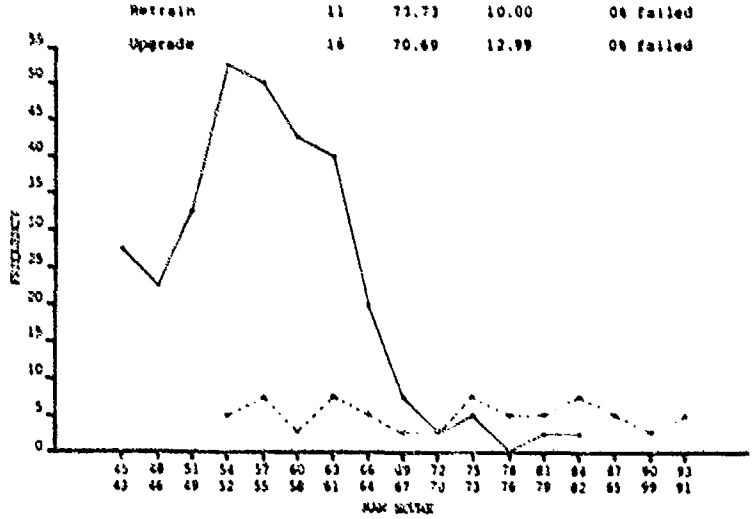
	n	Mean	S.D.	10th Percentile
Criterion Group	224	70.40	10.78	55
Examinee Group	71	49.27	13.21	
Bypass	53	43.09	11.63	88% failed
Retrain	30	53.53	12.90	53% failed



Two specialties, Plumbing and Engineering Assistant, showed high failure rates. Passing scores of 60 and 56 respectively were relatively high. Again, the examinee scores were highly varied. For the plumbing specialty, although the failure rate for the bypass group is the highest, the failure rate for the retraining group is also high. This suggests that those retraining may be coming from a variety of career fields and do not have the background knowledge required. For the engineering assistant specialty there is a much higher failure rate in the bypass group than in the retrain group. This suggests that knowledge required for this specialty may not be acquired in a civilian related job or there may not be a related civilian job. Again, the retrain group has a high failure rate that may suggest that those retraining are coming from a variety of career fields and lack the needed background knowledge. Also, the criterion group scores are higher than typical. Content of the test and training quality and emphasis may contribute to this effect.

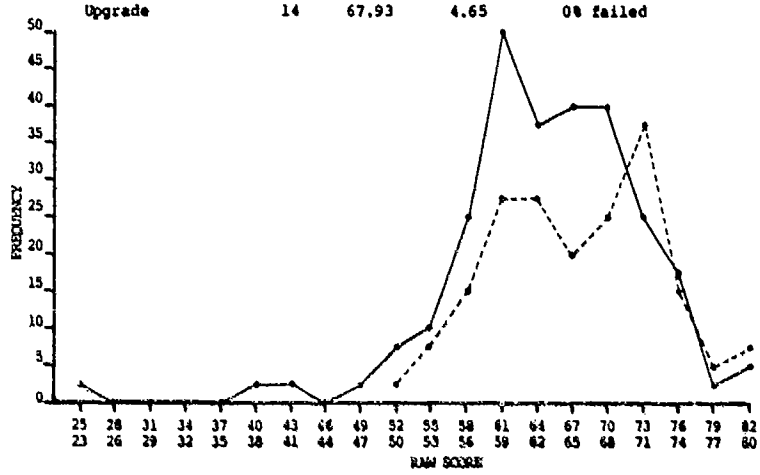
30630
APPRENTICE IMAGERY INTERPRETER SPECIALIST

	n	Mean	S.D.	10th Percentile
Criterion Group	123	55.97	7.59	45
Examinee Group	28	71.15	12.25	
Bypass	1	50.00		0% failed
Retrain	11	71.73	10.00	0% failed
Upgrade	16	70.69	12.99	0% failed



24130
APPRENTICE SAFETY SPECIALIST

	n	Mean	S.D.	10th Percentile
——— Criterion Group	108	63.79	8.23	56
----- Examinee Group	78	65.23	11.85	
Bypass	0	-	-	-
Retrain	64	64.64	12.85	9% failed
Upgrade	14	67.93	4.65	0% failed



AFSCs 20630, Apprentice Imagery Interpreter and 24130, Apprentice Safety Specialist, had AKTs with few or no failures. Both were different from the other AKTs studied in that only one test was administered for bypass and most tests were administered to Air Force Reserve and National Guard members either for retraining or upgrading purposes. Higher examinee means would be expected for these groups than for bypass groups. Airmen taking these exams may have already worked in the specialty or a very closely related specialty. In the case of the Safety specialty, some knowledge of that field is required for all specialties.

In general, all six AKTs exhibited two distinct characteristics. First, the variance in the distribution of scores was always greater for the examinee group than the criterion group. Though it can be expected that the criterion group, having just completed training in a specialty, would not vary much on a test covering that specialty, it was somewhat less expected that the examinee group scores vary so much more than the criterion group. For the Apprentice Medical Service Specialist test, the standard deviation for the examinee group was nearly five points greater than for the criterion group. Second, in looking at the subgroups of examinees, those taking the test for retraining and those for upgrading always had higher mean scores than those in basic training trying to bypass technical training. This result can be expected since those airmen retraining and testing for upgrading have been in the Air Force for a period of time already and have had an opportunity for more specific experience or, in the case of testing for upgrading, have been through on-the-job training in the specialty. These characteristics indicate that the criterion referenced tests are able to discriminate across varied groups of examinees.

Conclusions

For the AKTs analyzed, the higher means and relatively small standard deviations of the criterion groups provide a more precise pass/fail cutoff. It can be seen from the score distributions that when the scores at and below the tenth percentile for the criterion group are eliminated, the criterion group has greater homogeneity so that selection for bypass or retraining is similar to membership versus nonmembership in the criterion

group rather than achieving a specified criterion percentile across a distribution of criterion scores. This serves the expressed purpose of the AKTs to provide a means of selecting or not selecting an individual to bypass technical training.

For those AKTs with either very low or very high passing rates, criterion referenced tests were able to discriminate where the noncriterion referenced tests would have allowed too many or too few passing scores respectively.

Recommendations

Given large differences in mean scores of examinee and criterion groups, it is difficult to determine the validity of very high or very low passing rates. Performance studies of the bypass groups (Vaughan, 1978) should provide validity for the criterion cutoff scores. We are directing future research toward this goal.

Additionally, the high variance of examinee groups analyzed indicates a need for screening potential examinees. Given the wide range of examinee scores, some tests may be administered to airmen lacking the appropriate background knowledge or experience needed for a specialty. This suggests overuse of the tests and need for a better screening procedure.

References

Lord, F.M., and Novick, M.R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

Vaughan, D.S. Criterion-referencing the 47330 Apprentice Knowledge Test to technical school graduates. Lackland Air Force Base, Tx.: USAF Occupational Measurement Center Technical Note Series, March 1976. (No. 76-01)

Vaughan, D.S. Development of criterion-referenced Apprentice Knowledge Tests. Lackland Air Force Base, Tx.: USAF Occupational Measurement Center Technical Note Series, May 1976. (No. 76-04)

Vaughan, D.S. Job performance of USAF bypassed specialists. Oklahoma City, Oklahoma: proceedings of the 20th Annual Conference of the Military Testing Association, 1978.

Acknowledgements

The author gratefully acknowledges the following people for their assistance on this project: Mr. Paul Stanley, Chief, Occupational Test Development Branch; Lt Col Jimmy Mitchell; Maj Martin Dittmar; Capt Robert Worley; Ms. Marian Dike; Mr. Jim Barlow; SSgt Denise Born; Sra Anthony Jackson; A1C Bryan Hassenpflug; Mrs. Irma Bozart; Mrs. Velma Grantham; Ms. Anne O'Neill; and all the word processing section.