

113601  
~~15012~~

Resubmitted

GAO

United States General Accounting Office  
Washington, DC 20548

Logistics and  
Communications  
Division

B-200212

OCTOBER 6, 1980

The Honorable Harrison Schmitt  
United States Senate



113661

Dear Senator Schmitt:

Subject: [DOD's Carrier Evaluation and Reporting  
System] (LCD-81-6)

Your October 17, 1979, letter asked us to investigate the impact of the Department of Defense's (DOD's) Carrier Evaluation and Reporting System (CERS) on the moving and storage industry in New Mexico. As discussed with your office, prior staff commitments at that time delayed the start of our work until February 1980.

Soon after we began work on the assignment, it became apparent that many of the problems we observed, rather than being peculiar to agents in New Mexico, were inherent in CERS. Consequently, we expanded our review to include DOD installations in Colorado, Texas, and Arizona.

Although the CERS objective of high quality service at reasonable cost is good, current implementation practices preclude DOD's attaining this objective. We found that while administrative costs are increasing, additional benefits attributable to CERS are questionable. We also found that:

- DOD's emphasis on awarding shipments by cost over quality of service limits CERS' usefulness.
- The current evaluation process is unreliable.
- CERS' scores and rankings do not reflect quality of service.
- CERS' complexity causes confusion and diverse implementation.

(943350)

512293



--CERS does not appear workable in the peak shipping season.

We are recommending to the Secretary of Defense that a task force be established to determine whether the problems we have identified can be corrected or whether some other system of quality control should be instituted. Meanwhile, we are recommending that further expansion of CERS be deferred until the task force's findings are available.

During our review, we visited the following installations:


Holloman Air Force Base (AFB), New Mexico  
White Sands Missile Range,  
New Mexico  
Kirtland AFB, New Mexico  
Luke AFB, Arizona  
Fort Bliss Army Post, Texas  
Lowry AFB, Colorado  
Fort Carson Army Post, Colorado

Details of our findings, conclusions, and recommendations are described in enclosure I.

We forwarded advance copies of this report to the Secretary of Defense for comment on August 4, 1980. We did not receive a reply or a request for extension within 30 days. Therefore, in accordance with provisions of Public Law 96-226, this report does not include the views of agency officials.

As arranged with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from the date of the report. Then, we will send copies to interested parties and make copies available to others upon request.

Sincerely yours,

  
for R. W. Gutmann  
Director

Enclosures - 3

DOD'S CARRIER EVALUATION AND REPORTING SYSTEMBACKGROUND

As the manager of the Department of Defense's (DOD's) worldwide personal property moving and storage program, the Military Traffic Management Command (MTMC) is responsible for periodically evaluating the program's overall efficiency, economy, cost effectiveness, and adequacy. During the past 15 years, MTMC tried several times to develop systems to evaluate the overall performance of household goods carriers.

MTMC's first attempt at an automated data gathering system to evaluate carrier performance was the Worldwide Household Goods Information System for Traffic Management (WHIST). WHIST was established in 1966 and became operational in 1970. However, in 1974 we found many of the WHIST reports to be of little, if any, value because the information was incomplete, inaccurate, and untimely. WHIST's annual operating costs were about \$400,000 when it was discontinued as a quality control program in 1975.

MTMC's current carrier evaluation system (CERS) was developed in the mid-1970s, tested during 1976, and expanded nationwide in 1977. CERS is currently being used at 179 installations. MTMC is proposing to automate carrier evaluations nationwide (CERS II) and to expand CERS worldwide.

Before CERS, each installation developed an individual quality control program that implemented general MTMC policy guidance. Quality control was based on a system of actions (warnings, suspensions, nonuse, and disqualifications) for poor performance and each installation determined the particular emphasis of its own program.

Traffic was distributed to low-rate, qualified carriers on an equal-share basis, regardless of their levels of service quality. Carriers had no incentive to provide better service. As long as a carrier was rate-competitive and satisfactory, it shared equally in available traffic.

MTMC developed CERS to quantifiably measure the level of service that carriers were providing. To promote quality

improvements, it offered more shipments to higher performing carriers. CERS was also designed to standardize implementation of MTMC's quality control policies.

Under the CERS program, carrier performance on each shipment is graded from 0 to 100, with 100 representing perfect performance. Performance factors considered are ontime pickup and delivery, absence of loss or damage, customer satisfaction, and shipment handling and administrative procedures. A composite score is computed for each carrier, indicating its average performance for all shipments handled. Any carrier with a composite score below 50 cannot handle DOD shipments for at least 60 days.

Composite scores for all carriers serving a given shipping installation are arranged on a scale from high to low. The high 10 percent are considered "superior" carriers; the next 30 percent, "excellent;" and the other 60 percent, "standard."

Superior carriers receive twice the target tonnage and excellent carriers receive 1-1/2 times the target tonnage that standard carriers receive. Target tonnage is based on the volume of shipments expected at each installation.

In making allocations, both cost and performance are considered. Carriers with acceptable program ratings are stratified according to the rates they offered to DOD. At each rate level, the carrier's share depends on whether it is rated superior, excellent, or standard. At a given location, a standard carrier with a low rate would get tonnage after excellent or superior carriers at that rate level, but before higher performing carriers at higher rate levels. So, although performance is a key factor in tonnage distribution, rates are the first consideration.

#### COST INCREASED BUT NO EVIDENCE OF IMPROVED SERVICE

Initially, MTMC felt that CERS could be administered without additional personnel. However, most installations have added personnel to process CERS paperwork. These additional personnel costs are estimated to be \$3 million annually. In addition, MTMC and the Air Force have computerized some CERS data but could not estimate their developmental or operational costs.

Although costs have increased since its inception, we could not determine if service quality has improved as a result of CERS. MTMC claims of improved service were based on performance evaluation data. However, as described later, we found the CERS evaluative process to be unreliable because of current implementation practices.

CURRENT IMPLEMENTATION PRACTICES  
PRECLUDE ATTAINING CERS' OBJECTIVES

Numerous implementation practices preclude CERS from being a meaningful evaluation system. DOD's emphasis on cost over quality of service limits CERS' usefulness and often makes the system completely meaningless. Furthermore, CERS is based on assumptions, such as the existence of thorough inspection programs at each installation and the belief that individual shipment scores are meaningful representations of the quality of service on a particular move. However, these assumptions are not valid.

Present system design

CERS is designed to evaluate each shipment a carrier receives by both origin and destination shipping offices and by the service member. The shipping offices' evaluation is based on either an onsite inspection or a review of administrative documents, such as the bill of lading and inventory.

As noted previously, the carriers are graded on:

- Ontime pickup.
- Ontime delivery.
- Absence of loss or damage.
- Customer satisfaction.
- Compliance with the tender of service.

Point values are deducted based on the degree of importance assigned by MTMC. For example, 8 points are deducted for late delivery of 1 or 2 days; 40 points are deducted for late delivery of 10 or more days.

The individual shipment scores are averaged semiannually to find the overall carrier performance score for the period.

Emphasis on cost over quality

According to MTMC's quality control philosophy, "Quality control must be totally independent of cost. An unsatisfactory low cost shipment is just as bad as an unsatisfactory high cost shipment." Unfortunately, the current system does not consider rate levels in carrier evaluations, even though rates are the primary factor in shipment awards. Shipments are awarded to the lowest rate acceptable carrier, 1/ regardless of the carrier's performance quality level.

For example, a standard category carrier having the lowest rate at an installation will be offered as many shipments as it can manage. Unless more than one carrier is at that rate level, the CERS scores will never be used and the quality rankings are essentially ignored. Since rate level is the primary consideration in awarding shipments, some carriers tend to lower their rates to compete for shipments.

The Military Rate Tender is the standard rate DOD is willing to pay for shipping household goods. The Military Rate Tender is usually less than the standard commercial rate. To be rate-competitive, carriers must often lower their rates below the Military Rate Tender. Some carriers offer rates as much as 40 percent below the Military Rate Tender.

Rates below the Military Rate Tender were in effect year-round at five of the seven installations we visited. Installations with these reduced rates during the peak season usually were located on main north-south or east-west routes with extensive commercial traffic patterns. Having a reduced rate at these bases gives the carrier the option of moving DOD shipments if commercial traffic is not available. For example, one base awarded 99 percent of its traffic to low rate carriers during May to October 1979. However, it also had almost as many refusals as accepted shipments--800 refusals versus 1,000 accepted shipments. These carriers were not only receiving large numbers of shipments, but they were also refusing an equal number of shipments. One carrier accepted 132 shipments, while refusing 120.

---

1/An acceptable carrier has an overall carrier performance score of at least 50.

The installations we visited awarded 80 to 99 percent of all shipments annually to carriers with rates below the Military Rate Tender. During the peak 1979 season (May to Oct.), about 93 percent of the installations' shipments were awarded to these carriers.

Because more shipments are awarded to higher ranked carriers than lower ranked carriers (within a given rate level), CERS appears to foster rate reductions by low-ranked carriers. For example, at four of the five installations where reduced rates were in effect, standard carriers filed these rates for the peak season and received a substantial number of shipments. Without a reduced rate, these carriers probably would not have received as many shipments.

As shown below, at one installation, a standard carrier received the second largest number of shipments after filing a reduced rate effective from May to August 1979 for 20 States.

<u>Carrier</u>	<u>Ranking as of May 1979</u>	<u>No. shipments tendered (May- Oct. 1979)</u>
1	Excellent (8 of 48)	99
2	Standard (30 of 48)	74
3	Excellent (18 of 48)	66
4	Standard (22 of 48)	43
5	Superior (2 of 48)	39

Although the excellent carrier's rate was higher than the No. 2 standard carrier's, the rate was in effect for 45 States. This enabled the higher priced excellent carrier to receive more shipments than the lower priced standard carrier.

Another example of the apparent fostering of reduced rates recently occurred at an installation where, previously, all carriers were at the Military Rate Tender level. The carrier ranked 16th of 22 filed a reduced rate for the 1980 peak season. The agent for this standard carrier said that the rate was probably filed because the carrier would not have received many shipments due to its low ranking. Reducing

the rate was the only way this standard carrier could receive a large number of shipments from the installation. As currently structured, low ranking carriers can file reduced rates to receive significant amounts of tonnage, thus bypassing the entire CERS ranking system.

Shipment evaluations are unreliable

Although CERS assumes that each installation will develop and execute a thorough inspection program, this has not occurred. The lack of viable inspection programs at each installation makes CERS evaluations unreliable and of questionable value. Without thorough inspection programs, the shipment scores do not reflect actual carrier performance or measure the true quality of carrier service. CERS' evaluation deficiencies include

- incomplete shipment evaluations, which inflate scores;
- statistically biased selection of shipments for inspection;
- limited quantity and quality of inspections; and
- inadequate member evaluations.

Incomplete evaluations inflate scores

CERS assumes that each shipment will receive three evaluations: at both origin and destination (by either an onsite inspection or a review of administrative documents) and by the service member. In reality, many shipments are not evaluated or receive only partial evaluations. At several installations, only 10 to 20 percent of all shipments received all three possible evaluations. Shipping offices seldom used administrative documents in place of onsite inspections. Since any part of a shipment that is not evaluated receives full credit points, many shipments receive inflated scores. The worst example is a completely unevaluated shipment which received a perfect score of 100.

The lack of complete shipment evaluations, combined with CERS procedures that fully credit unevaluated shipments, causes inflated scores, severely limiting their reflection of actual service quality.



Inspection selection is statistically biased

CERS assumes that each installation will develop a performance data base from shipment evaluations and scores. This data base should represent actual carrier performance. For the inspected shipments to statistically represent all shipments tendered, inspections should be selected randomly. However, data bases are biased by the following installation practices which preclude random selection.

- Limiting the distance inspectors travel. For example, one installation is responsible for shipments within three States, but its inspectors were limited to 25 miles of travel per day.
- Concentrating on a number of shipments in the same geographic area. For example, inspectors at one installation inspected a large number of shipments located close to one another instead of a few individual shipments dispersed over a wide area.
- Concentrating on large, high-tonnage shipments.
- Concentrating on shipments handled by lower ranked carriers.
- Giving some shipments preferential treatment. For example, inspectors at one installation said moves involving high ranking military members receive more thorough inspections than those involving low ranking members.

Nonrandom selection of inspected shipments means CERS scores may not accurately reflect actual carrier performance.

Limited quantity and quality of inspections

Shortages of installation inspection personnel and administrative restrictions, such as travel limitations, have adversely affected the number and quality of shipment inspections. One Air Force audit reported installation inspections were mostly "paper exercises" because of personnel limitations. We agree. At one installation visited, we found that one inspector managed the entire quality control program. Instead of the 50-percent DOD inspection requirement, the inspector reviewed only about 15 percent of all

personal property shipments--only some of which were CERS shipments. The inspectors are also responsible for inspecting shipments to and from overseas, unaccompanied baggage, mobile homes, and local moves.

At another installation with four inspectors, only about 20 percent of all origin and destination shipments were inspected--many only cursorily. Some of these inspections consisted of merely observing 5 or 10 percent of the packing. Since carriers receive full credit for any part of a shipment that is not evaluated, restricted inspection programs limit the relationship between scores and actual carrier performance.

#### Member evaluations are inadequate

The lack of adequate inspections places a greater reliance on the service member to evaluate the carrier's performance. However, the return rate for these evaluations is low, and many returned evaluations are either ignored or changed. A nonevaluation by a member may indicate satisfaction with the move or it could also indicate a lack of interest with the program.

MTMC has not set a standard for an acceptable return rate, but has termed return rates below 55 percent as unacceptable. However, one Air Force study at 23 installations showed a 40-percent return rate. Return rates at the installations we visited were as low as 30 percent; at only one installation was it above 50 percent. Even when members did evaluate a carrier, shipping offices often changed negative comments. At some installations, one-third to one-half of all member evaluations with negative comments were changed or ignored because the comments did not concern contractual violations and/or did not contain enough detail to satisfy the shipping office.

For example, one member reported being dissatisfied with the mover at his old residence because of a late pick-up and an estimated \$100 to \$200 in damages--a potential 60 points in penalties. The base shipping officer did not feel that the member had supplied enough detail, so he assessed no penalty points. The carrier received a perfect score of 100.

Carrier performance assessment is critically dependent on member evaluations, particularly when inspections are

nonexistent or incomplete. However, few members complete their evaluation, and even then, penalty points may not be assessed because members' comments are ignored or changed by the shipping offices. This results in inflated scores, again precluding an accurate reflection of service quality.

CERS scores and rankings do not reflect quality of service

CERS assumes that carrier scores and rankings reflect measurable differences in service quality. However, numerous deficiencies in the scoring and ranking processes prevent the fulfillment of this assumption. These deficiencies include

- arbitrary performance standards,
- meaningless percentage rankings,
- incompatible suspension and reinstatement practices, and
- a lack of performance history requirements.

Arbitrary performance standards

In developing CERS performance standards, MTMC established

- an average score of 50 as the minimum acceptable performance and
- the 10/30/60 percentage ranking system, by which carriers scoring 50 or above are ranked as superior, excellent, or standard, respectively.

However, MTMC cannot relate these standards to actual carrier performance. For example, MTMC does not know if 50 is too high, too low, or an adequate minimum performance level. Realistic standards are essential to adequately evaluate carrier performance.

10/30/60 categories not meaningful

At most installations we visited, small point differences separated carriers in the superior, excellent, and standard categories. For example:

- At one installation, 6 of 11 carriers scored between 90 and 100; yet because of the 10/30/60 rule, only 1 could be ranked superior, 3 excellent, and 2 standard. The difference between the standard and excellent categories was less than two points.
- At another installation, 10 of 22 carriers scored between 90 and 100, but only 2 could be ranked as superior, 7 excellent, and 1 standard. The difference between the superior and excellent categories was less than three-tenths of a point, and the difference between the excellent and standard categories was about one point.
- Another installation had 34 of 73 carriers scoring between 80 and 89. Seven were ranked superior, 22 excellent, and 5 standard. The difference between the superior and excellent categories was four-tenths of a point, and only one-tenth of a point separated the excellent and standard categories.

We also found that if one eliminates penalty points for shipment refusals--that is, considers only the scores on shipments actually moved--there is no statistically significant difference in the scores. The scores of shipments actually moved for selected carriers at two installations were tested for statistical differences. At one installation, the No. 1 rated superior carrier was tested against the No. 22 rated excellent carrier and the No. 49 rated standard carrier. We concluded that no quantifiable difference existed between the score of the No. 1 carrier and the scores of the other carriers. At another installation, the No. 2 superior carrier was tested against the No. 3 excellent carrier, with the same result. Highlights of our test are shown in the following table. A detailed analysis is included as enclosure II.

Comparison of Composite Performance Scores

Carrier ranking	Installation A		Carrier ranking	Installation B	
	HTMC's scores	GAO's revised scores		HTMC's scores	GAO's revised scores
1-Superior	89.63	95.25	2-Superior	94.61	97.42
22-Excellent	83.00	93.51	3-Excellent	94.39	96.85
49-Standard	75.33	91.12	-	-	-

Performance histories  
not required

CERS does not require a performance history before allowing new carriers to advance in the ranking. Rather than considering consistent and sustained service quality, the current system allows carriers to be ranked superior based on only one shipment. At almost every installation we visited, carriers were advancing from standard to superior rankings based on very few shipments that were either partially evaluated or not evaluated at all. For example, during one period, a new carrier with an administrative score 1/ of 75 was ranked 31st out of 34 carriers. Then, after receiving a score of 100 on an unevaluated shipment, this carrier was ranked number one (superior) the following period and received more shipments than any other carrier.

At another installation, a carrier became superior (2d of 37) based on two partially evaluated shipments. During the next period, the carrier dropped to excellent (14th of 38) based on four other shipment scores. Enclosure III provides more examples of such fluctuations.

Without an adequate historical performance record, MTMC cannot determine a carrier's overall quality.

CERS' COMPLEXITY CAUSES CONFUSION  
AND DIVERSE IMPLEMENTATION

A 1979 Air Force report said that CERS creates an administrative burden and causes confusion. Although designed to standardize carrier performance evaluations and shipment awards, CERS is ineffective in doing so because of its diverse and inaccurate implementation.

Inconsistencies and inaccuracies  
in carrier evaluations

According to MTMC, carrier evaluations must be objective and standardized, but in practice, installation evaluation and scoring processes are subjective and varied, as illustrated by the following table of procedures used at the installations we visited.

---

1/A score is administratively assigned to new carriers because a performance history is lacking.

Inconsistencies in Evaluating and Scoring Carrier Performance

<u>Inconsistencies in penalty point assessment</u>	<u>Installation</u>						
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>G</u>
<b>Loss and damage claims:</b>							
Usually accept member's allegation at face value	X	X	X	X			X
Require additional documentation					X	X	
<b>Shipments from storage:</b>							
Shipments generally not scored					X		
Loss and damages ignored	X						
Shipment scored and up to the carrier to refute		X	X	X		X	X
Origin scores not counted			X	X			
<b>Customer satisfaction reports:</b>							
Accept at face value without much analysis			X				
Require detailed support related to tender of service		X		X	X	X	X
Use judgment, not necessarily related to tender of service	X						
<b>Refusals to accept shipments:</b>							
Usually penalize refusal		X	X	X	X		X
Penalize only if no other carrier will accept shipment	X						
Not consistently penalized						X	
<b>Other:</b>							
Never deduct points for "unqualified personnel"	X						
Ignore customer dissatisfaction if form not signed	X						
Ignore adverse information on documents other than the MTMC forms (such as letters or other review forms)	X						
Ignore improper packing points unless a loss/damage estimate is given				X			

The 1979 Air Force audit report identified similar inconsistencies in carrier evaluations and scoring, including:

--Shipment refusals not penalized.

--All evaluations not considered.

Such inconsistencies caused incorrect scoring and ranking of many carriers. Correct scores would have placed some carriers in different performance ranges.

Inconsistencies and inaccuracies  
in tonnage allocation

CERS is designed to reward superior carrier performance (as determined by evaluation and scoring) with increased, or "reward," tonnage. However, as we have demonstrated, the evaluative processes are not reliable. Even if carrier performance could be consistently and accurately measured, inconsistencies and inaccuracies in tonnage allocation practices would still preclude the reward system from working as designed. Installation personnel attributed the inconsistencies and inaccuracies to complex and confusing CERS procedures.

Tonnage allocation practices vary considerably among installations. For example, some installations award shipments alphabetically (within a performance range), rather than by performance score. Another installation incorrectly saturated the top ranking carrier with tonnage, instead of spreading it proportionately among performance ranges. This, in effect, excluded the other qualifying carriers from receiving their fair shares of tonnage.

Diverse reinstatement procedures also inhibit consistent tonnage allocation. Reinstatement procedures at four of the seven installations were different than those contained in the CERS manual. For example, one CERS procedure for reinstating a suspended carrier is that the carrier will be penalized the highest amount of tonnage that any other carrier received during the suspension. We found that some bases were ignoring this procedure.

Since the prime consideration in tonnage allocation is rate level, carrier rate information must be correct, complete, and current. If it is not, tonnage cannot be correctly and equitably allocated. Having correct rate

information is especially critical for CERS, which awards incentive tonnages.

For example, if a superior carrier is incorrectly listed at the wrong rate level, it may receive incentive tonnage over a standard carrier correctly listed at that rate level.

The 1979 Air Force audit report noted inaccuracies in rate level data at almost two-thirds of the installations it reviewed. Examples of such rate data inaccuracies include carriers listed

- without rates on file,
- with expired rates,
- at the wrong rate level, and
- for the wrong States.

The report concluded that failure to maintain accurate carrier rate data may result in the award of traffic to other than the appropriate carrier.

#### CERS HAS NOT WORKED IN THE PEAK SEASON

Quality control is most important during the peak moving season--from May to October. During this period, the military must compete with the civilian sector for limited carrier resources. This competition is difficult for DOD because military shipment rates are lower than those of civilian shipments.

Service quality has historically deteriorated during the peak season. During the summer of 1973, significantly deteriorating service prompted MTMC to develop the CERS quality control program. The program's intent was to provide high quality moves year-round. Although CERS was implemented nationwide in 1977, peak season problems have continued to plague DOD. Truckers' strikes and fuel shortages in recent summers have aggravated peak season problems, but the problems continued even after the strikes and shortages ended.

We could not find any discernible service quality improvements that can be credited to CERS. Several DOD



organizations, for example, termed the 1978 peak moving season a disaster. One Air Force headquarters reported new highs in missed delivery dates and inconvenienced service members.

Problems again occurred during the 1979 peak season, when some installations had almost as many shipments refused as accepted. As shown in the following examples, some shipping offices ignored or suspended CERS procedures and awarded shipments to any carrier willing to accept them.

- Not properly assessing penalty points for shipment refusals. This invalidates CERS scores and makes accurate comparison of carrier performance impossible.
- Reinstating suspended carriers early, without assigning penalties. This negates the system's penalty provisions.
- Suspending use of the tonnage distribution roster and awarding shipments to whomever will accept them. At one installation, with all carriers at the Military Rate Tender, standard carriers received 47 percent of the peak season shipments, while superior carriers received only 11 percent. This negates CERS' incentive concept.
- Not consistently assessing penalty points for missing delivery dates. This prevents uniform application of CERS' basic standards and scoring.

Peak season performance is the most crucial test of any carrier evaluation system. Any such system must be able to evaluate carrier performance during the period when the largest number of moves occurs. If an evaluation system is unmanageable or ineffective during the peak season, then the value of the system is questionable.

Because it fiercely competes with the civilian sector for limited carrier resources, DOD finds it difficult to provide adequate incentive for high quality moves during the peak season. As one Air Force report noted, the low rates paid by DOD contribute to poor service because some carriers decline Government shipments in favor of private shipments (which offer greater revenue).

CONCLUSIONS AND RECOMMENDATIONS

Although CERS' administrative costs have increased by almost \$3 million annually, we could not determine if service quality has improved because of CERS. The current CERS program has numerous problems and is often ignored during the peak season when quality control is most important.

This is the second major quality control system developed by MTMC over the last 15 years. Millions of dollars have been spent to administer and develop these systems, but neither has accomplished its intended objectives. The complexity and subjectivity of data gathering at each installation, coupled with peak season and rate problems, make it extremely difficult to develop such a system.

We are recommending to the Secretary of Defense that a task force be established to review MTMC's quality control requirements, the resources available to meet these needs, and the potential for correcting the problems identified in this report. We also recommend that MTMC plans for expanding CERS into phase II and worldwide be postponed until the task force completes its review.

DETAILED ANALYSIS OF CERS RANKINGS

Hypothesis testing is a statistical procedure used to verify or nullify preconceptions about population parameters. We used this method to statistically test for significant differences in carrier scores used by DOD in assessing quality of service rendered by the carriers. Our tests at two installations showed that the differences in scores used are not statistically significant and do not reflect a quantifiable difference in quality.

ONE-TAILED STUDENT'S T-TEST USED TO TEST FOR SIGNIFICANCE

We calculated average scores from the raw scores of selected carriers per period. Raw scores are the actual shipment scores. We did not use refusal points in our calculations since they were inconsistently applied or ignored and could not be measured for statistical significance. We expanded DOD's method of calculating new composite performance scores--multiplying the old score by 0.6 and the new average shipment score by 0.4 and adding the results--to derive appropriate weights per period at the two installations. We then used these average scores and weights to calculate a weighted average shipment score for the selected carriers. Weighted variances were similarly calculated for the selected carriers.

Hypothesis testing was performed on the weighted average scores, the null hypothesis being that a higher rated carrier's weighted score is equal to a lower rated carrier's weighted score, and the alternative hypothesis being that a higher rated carrier's weighted score is greater than a lower rated carrier's weighted score. Mathematically:

$$H_0: u_h - u_l$$

$$H_1: u_h > u_l$$

where:  $H_0$  = the null hypothesis

$H_1$  = the alternative hypothesis

$u_h$  = the higher rated carrier's weighted score

$u_l$  = the lower rated carrier's weighted score

A one-tailed student's t-test was used for this hypothesis testing.

TEST RESULTS OF DOD DATA SHOWED NO  
DIFFERENCE IN CARRIER SCORES

Results of the t-tests showed that the differences in scores are not statistically significant and do not reflect quantifiable differences in quality. At one installation, the No. 1 rated superior carrier was tested against the No. 22 rated excellent carrier and the No. 49 rated standard carrier. At another installation, the No. 2 superior carrier was tested against the No. 3 excellent carrier. In all of these tests, the calculated t-value was less than the critical t-value from the student's t-distribution, at a 95-percent confidence level. Accordingly, the null hypothesis, that a higher rated carrier's weighted score is equal to a lower rated carrier's weighted score, was not rejected in every case. Therefore, by not rejecting the null hypothesis, we concluded there was no difference in quality of service rendered by the carriers.

The following table shows the carriers' ranking and scores at two installations.

Carrier Scores And Rankings As Analyzed By GAOInstallation A

<u>Carrier</u>	<u>Score</u>	<u>Carrier performance ranking</u>	<u>Carrier</u>	<u>Score</u>	<u>Carrier performance ranking</u>
1	89.63	Superior	44	76.41	Standard
2	89.12	"	45	76.31	"
3	88.70	"	46	76.23	"
4	88.13	"	47	76.04	"
5	87.82	"	48	75.74	"
6	87.66	"	49	75.33	"
7	87.62	"	50	75.32	"
			51	74.76	"
8	87.21	Excellent	52	74.02	"
9	86.17	"	53	73.48	"
10	84.77	"	54	73.43	"
11	84.49	"	55	72.53	"
12	84.48	"	56	72.52	"
13	84.34	"	57	71.34	"
14	84.18	"	58	71.10	"
15	84.13	"	59	68.90	"
16	84.00	"	60	68.51	"
17	83.98	"	61	65.98	"
18	83.35	"	62	64.00	"
19	83.31	"	63	63.71	"
20	83.17	"	64	63.52	"
21	83.09	"	65	62.71	"
22	83.00	"	66	61.84	"
23	82.60	"	67	60.50	"
24	82.66	"	68	60.00	"
25	82.32	"	69	54.86	"
26	82.09	"	70	50.40	"
27	81.98	"	71	50.00	"
28	81.52	"	72	50.00	"
29	81.50	"	73	50.00	"
30	81.44	Standard			
31	81.05	"			
32	80.85	"			
33	80.48	"			
34	80.09	"			
35	79.00	"			
36	77.84	"			
37	77.68	"			
38	77.64	"			
39	77.59	"			
40	77.534	"			
41	77.53	"			
42	76.54	"			
43	76.44	"			

Installation B

<u>Carrier</u>	<u>Score</u>	<u>Carrier performance ranking</u>
1	100.0	Superior
2	94.613	"
3	94.396	Excellent
4	94.116	"
5	93.923	"
6	93.178	"
7	91.75	"
8	91.214	"
9	91.212	"
10	90.114	Standard
11	88.0	"
12	87.062	"
13	83.08	"
14	81.52	"
15	81.0	"
16	79.965	"
17	77.478	"
18	74.8	"
19	59.26	"
20	50.0	"
21	50.0	"
22	50.0	"

EXAMPLES OF CARRIER RANKING FLUCTUATIONS

<u>Instal- lation</u>	<u>Previous score</u>	<u>Previous ranking</u>	<u>New score</u>	<u>New ranking</u>	<u>No. of shipments scored</u>	<u>Explanation</u>
A	a/75 (as of 3-31-78)	Standard (23 of 29)	97.14 (as of 11-1-78)	Superior (1 of 34)	1	Became No. 1 based on one score
B	None (as of 11-1-78)	Standard (42 of 45)	100 (as of 5-1-79)	Superior (1 of 37)	1	Became No. 1 based on one score
C	100 (as of 5-1-79)	Superior (1 of 47)	100 (as of 11-1-79)	Superior (1 of 48)	0	Retained No. 1 ranking based on no scores
D	100 (as of 5-1-79)	Superior (2 of 75)	100 (as of 11-1-79)	Superior (1 of 62)	0	Advanced from No. 2 to No. 1 based on no scores
E	a/50 (as of 11-1-79)	Standard (21 of 23)	100 (as of 5-1-80)	Superior (1 of 22)	1	Became No. 1 based on one score
F	81.86 (as of 5-1-79)	Standard (50 of 83)	89.12 (as of 11-1-79)	Superior (2 of 73)	2	Became No. 2 based on two scores
G	50 (as of 5-1-79)	Standard (85 of 89)	100 (as of 11-1-79)	Superior (1 of 91)	2	Became No. 1 based on two scores

a/Score was administratively assigned because carrier was new and had no performance history.

