January 31, 2000

A. KEITH STRANGE
VICE PRESIDENT, PURCHASING AND MATERIALS

SYLVESTER S. BLACK
VICE PRESIDENT, NETWORK OPERATIONS MANAGEMENT

SUBJECT: Commercial Air Carrier Performance: Performance Measurement System (Report Number TR-AR-00-003)

This is the first of three reports that presents the results of our audit of commercial air carrier performance. Our audit disclosed that the performance measurement system does not provide the Postal Service and commercial air carriers with the type of information needed for performance improvements. Also, mail volumes measured are not representative of overall air carrier performance and data supporting performance measurement is inaccurate. We recommended that the use of the current performance measurement system be suspended, and either redesign the performance measurement or explore other alternatives for improving air carrier performance.

Management agreed with the report findings and recommendations. Based on our report and a series of issues the Postal Service has been working on, the Postal Service will suspend use of the present performance measurement system on February 26, 2000. They will also explore other options for improving air carrier performance. Management's comments and our evaluation of these comments are included in the report.

We appreciate the cooperation and courtesies provided by your staff during the audit. If you have questions or need additional information, please contact [redacted], director, Transportation, at [redacted], or me at (703) 248-2300.

Richard F. Chambers
Assistant Inspector General
for Performance

Attachment

cc: Clarence E. Lewis, Jr.
John E. Potter
Tony M. Pajunas
John R. Gunnels

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Introduction

This is the first of three reports resulting from our audit of commercial air carrier performance.[1] The chief operating officer requested that we review air carrier performance because carriers have not been meeting expected on-time delivery targets established by the Postal Service's Air Systems contract. In fiscal year (FY) 1999, the national average score for on-time performance was 60 percent, as reported by the Postal Service.

This report presents our assessment of the reliability of the performance measurement system used to evaluate air carrier performance. Specifically, it addresses whether (1) the performance measurement system provides the Postal Service and air carriers with the type of information needed for performance improvements, (2) performance scores are representative of air carrier performance, and (3) data supporting performance measurement is accurate.

Results in Brief

The performance measurement system does not provide the Postal Service and commercial air carriers with the type of information needed for performance improvements. Although the system scores on-time delivery, it does not identify the causes for late mail arrival or measure all activities that impact air carrier performance. As a result, the Postal Service and air carriers cannot take corrective action, and air carriers may receive low scores for delays outside of their control. Further, because the system measures only mail that successfully arrives at its destination on its intended day of arrival, it does not consider late mail that is rolled over to the next day for delivery. The exclusion of this mail may cause carriers to receive more favorable performance scores, although mail has missed its intended flight.

Because the performance measurement system tests only Priority Mail, which comprises less than 34 percent of total mail volumes commercially flown, performance scores are not representative of overall air carrier performance. Further, the volume of Priority Mail tested is insufficient because scan rates are below minimum levels established for statistical reliability at some sites. Low scan rates are attributable to inadequate equipment and staff assigned to

[1] The second report will address the effectiveness of the Air Systems contract in improving air carrier performance and the third report will address ground handling services performed by postal employees.

scanning operations, and direct routing of mail to Priority Mail processing centers that bypass scanning sites. As a result, air carriers and postal managers lack confidence the performance measurement system reliably assesses air carrier performance.

Our review also disclosed that data supporting performance measurement is inaccurate. Specifically (1) employees responsible for scanning operations did not always reset clocks to capture actual delivery times, and (2) scan rates for testing on-time delivery were inaccurate due to a reliance on historical versus actual mail volumes and the inclusion of air contract tagged mail. Without accurate data the Postal Service cannot hold carriers accountable for their performance or initiate corrective actions needed to improve on-time delivery.

Summary of Recommendations

We recommend the chief operating officer suspend use of the current performance measurement system, and either redesign performance measurement or explore other alternatives for improving air carrier performance.

Summary of Management's Comments

Management agreed with our findings and recommendations. Based on our report and a series of issues the Postal Service has been working on, the current performance measurement system will be suspended on February 26, 2000. Management also agreed to explore other options for improving air carrier performance. Management's comments are included in their entirety in Appendix B of this report.

Evaluation of Management's Comments

Management's comments are responsive to our findings and recommendations, and planned actions should identify a more effective alternative for improving air carrier performance.

INTRODUCTION

Background

In FY 1999 the Postal Service spent over $670 million under its Air Systems contract with air carriers to move more than 2.1 billion pounds of Priority Mail and First-Class Mail. Although the Air Systems contract established a 98 percent on-time delivery target, air carrier performance averaged 60 percent in FY 1999, as reported by the Postal Service. Because air carrier performance directly impacts on-time delivery, air carrier delays can hinder the Postal Service's ability to achieve service commitments for these classes of mail.

In 1994 the Postal Service developed a performance measurement system to identify and track mail from the time it is scanned by the Postal Service at the origin processing facility until it is delivered by the air carrier to the destination airport mail center/facility. The purpose of the system is to assess air carrier performance in achieving on-time delivery goals and provide information required for corrective action. However, both the Postal Service and air carriers have expressed concern about the system's ability to accurately measure performance.

To assess air carrier performance, mail is tested for on-time delivery at 79 centers/facilities that receive the largest volumes of incoming Priority Mail and First-Class Mail. Currently the system tests only Priority Mail pieces based on random sampling. Tracking selected pieces of mail is accomplished by using the Scan-Where-You-Band system. This system, used at the origin facility, scans a barcoded distribution label, and produces a dispatch and routing tag. At the destination airport mail center/facility, hand-held or fixed scanners are used to read the tags of inbound test pieces to determine if the mail was delivered to the Postal Service by the scheduled delivery time.

Performance scores are generated from the scanned data and used to establish air carrier rankings. These rankings allow the Postal Service to group airlines in specific origin/destination market pairs based on delivery performance. Quarterly performance rankings are used to shift mail volumes tendered to air carriers in higher performance groups. The air carrier with the highest score for a particular market is the preferred carrier for that market pair.

Objective, Scope, and Methodology

The objective of our audit was to assess the reliability of the system used to measure air carrier performance. To accomplish this objective, we determined whether (1) the performance measurement system provides the Postal Service and air carriers with the type of information needed for delivery improvements, (2) performance scores are representative of air carrier performance, and (3) data supporting performance measurement is accurate.

To determine whether the system provides the Postal Service and air carriers with the type of information needed for performance improvements, we reviewed the Postal Service's Performance Measurement Handbook, and its sampling methodology and procedures for testing mail pieces. We also interviewed postal logistics staff in headquarters responsible for managing the system, performance measurement coordinators and specialists in the field, and the statistician who developed the performance measurement sampling and test procedures.

To determine whether performance scores are representative of air carrier performance and supporting data is accurate, we visited the following ten airport mail centers/facilities that had the largest volume of scanned Priority Mail during the first three quarters of FY 1999:[2]

Site	Average Volume of Scanned Mail
Boston	107,533
Honolulu	67,609
Chicago	58,243
Philadelphia	54,371
Dallas	53,386
Orlando	51,161
Kansas City	39,979
Baltimore	37,857
Milwaukee	37,290
New York LaGuardia	36,261

At these sites we observed scanning operations, reviewed tender and delivery procedures, and interviewed postal distribution network office managers, performance measurement coordinators, plant mangers, and network specialists. We also reviewed National Traffic Management

[2] The first three quarters of FY 1999 covers the period, September 12, 1998, to May 21, 1999.

System reports, local databases and spreadsheets, and other relevant documents. We reviewed policies and procedures, the Air Systems contract, headquarters directives, and met with the appropriate headquarters officials. We also interviewed representatives of six major commercial air carriers[3] to obtain their insight regarding the performance measurement system.

This audit was conducted from May 1999 through December 1999 in accordance with generally accepted government auditing standards and included tests of internal controls as were considered necessary under the circumstances. We discussed our findings with appropriate management officials and included their comments, where appropriate.

Prior Audit Coverage

One Postal Service and two Postal Inspection Service reports[4] issued during the last five years addressed several areas where management needed to strengthen the performance measurement system. These included identifying the causes of air transportation delays, and consistently applying scanning procedures. Management agreed with these findings, but had not implemented the Inspection Service recommendation to include First-Class Mail in performance measurement.

[3] We interviewed representatives from American, Continental, Northwest, Transworld, USAir, and United airlines.
[4] Developing a 21st Century Air Transportation Strategy for the Postal Service, (March 17, 1999); National Coordination Audit: Air Transportation Network, Case Number 023-1209811-PA(1), (May 1998); and National Review of Airport Mail Centers, Case Number 024-1165776-PA(1), (November 1995).

AUDIT RESULTS

Performance Measurement Design

The performance measurement system does not provide the Postal Service and commercial air carriers with the type of information needed for corrective action. Although the system scores on-time delivery, it does not identify the causes for late mail arrival or measure all activities that impact air carrier performance. For example, the system does not identify whether delays are due to the Postal Service over-assigning mail to flights or tendering the mail late, or to carriers not loading mail onto designated flights. As a result, the Postal Service and air carriers cannot take corrective action, and air carriers may receive low scores for delays that are not always within their control.

Further, because the performance measurement system measures only mail that successfully arrives at its destination on its intended day of arrival, it does not consider late mail that is rolled over to the next day for delivery. As a result, air carriers may receive more favorable performance scores, although mail has missed its intended flight and is delivered the following day.

These issues were first identified in a March 1999 study conducted by a consulting firm under contract with the Postal Service.[5] This study concluded that the performance measurement system did not identify the causes of air transportation delays, making it impossible to determine what or who is at fault and how deficiencies can be corrected. The study further noted that due to these problems, the Postal Service and air carriers lacked confidence that the system provided information required for improving performance.

Further, a 1998 Baldridge study[6] of postal operations concluded the Postal Service needed to address the on-time performance of air carriers through process management.[7] In response to this report, the Postal Service is implementing process management to improve operations. However, because the performance

[5] Developing a 21st Century Air Transportation Strategy for the Postal Service, (March 17, 1999).
[6] Assessing USPS' Management System - Feedback Report, (May 27, 1998).
[7] Process management is the documentation of an organization's core processes and that of its suppliers in an effort to better manage and improve performance.

measurement system for air carriers does not provide the information needed for process management, the Postal Service will need to identify and measure all aspects of carrier performance to improve air carrier contributions to Postal Service performance goals.

Representative Scores	Because the performance measurement system tests only Priority Mail, which comprises less than 34 percent of total mail volumes commercially flown, performance scores are not representative of overall air carrier performance. Further, the volume of Priority Mail tested is insufficient because scan rates are below minimum levels established for statistical reliability at some sites.
Mail Classes Measured	Although commercial air carriers transport primarily Priority Mail and First-Class Mail, only Priority Mail is measured. Priority Mail comprises only 34 percent of all mail commercially flown. Postal officials stated First-Class Mail, which accounts for 55 percent of mail transported by commercial air carriers, is not tracked under performance measurement because it is not a competitive product. However, managers at eight[8] of the ten sites we visited believed including First-Class Mail in performance measurement would provide a more representative assessment of air carrier performance. In a May 1998 report on the Postal Service's air transportation network,[9] the Postal Inspection Service recommended the performance measurement system be enhanced to include tests of First-Class Mail. The Postal Service agreed with the findings, but stated the lack of equipment and additional man-hours required for testing First-Class Mail were barriers to implementing the Inspection Service's recommendation. The Postal Service agreed to address these barriers and expected to mandate scanning of First-Class Mail by February 1999. The Postal Service had not implemented this recommendation at the time of our review; however, officials are exploring software and hardware changes needed for the enhancement. Although the Postal Service's reasons for restricting performance measurement to Priority Mail may have been appropriate at the time, air carrier performance directly impacts the delivery of First-Class Mail, which comprises 59 percent of postal revenues. With expected diversion of First-Class Mail volumes to the Internet and electronic alternatives, further losses due to poor performance may significantly impact Postal Service revenues. For this

[8] Managers at the Baltimore, Boston, Chicago, Dallas, Kansas City, Milwaukee, New York La Guardia and Orlando Airport Mail Centers/Facilities believed that including First-Class Mail would provide a more representative assessment of air carrier performance.

[9] National Coordination Audit: Air Transportation Network, Case Number 023-1209811-PA(1), (May 1998).

reason, the Postal Service must measure air carrier performance to improve the delivery of First-Class Mail.

Volume Scanned	For performance measurement, the Postal Service requires test sites scan a minimum of 67 percent of incoming mail. Despite this requirement, scan rates at 17 of 79 sites were consistently below this level during the first three quarters of FY 1999.[10] Of the 17 sites that had low scan rates: • Five did not scan at all. • Three never reached the 67 percent level. • Nine achieved the 67 percent scan rate less than 25 percent of the time. Scan rates reported for the 79 sites are provided in Appendix A. Postal Service officials attributed low scan rates to inadequate equipment and staff assigned to scanning operations, direct routing of mail to Priority Mail Processing Centers that bypass scanning locations, and use of air contract tags that cannot be scanned. These causes were also confirmed by our review. Three sites[11] were excluded from performance measurement because they lacked adequate equipment for scanning high mail volumes. Site visits disclosed four[12] of ten locations relied on hand-held scanners that could not scan all of the mail, and at another site[13] two of four fixed scanners were inoperable. Further, inefficient induction systems at three[14] sites limited the number of mail pieces that postal employees could scan.

[10] The first three quarters of FY 1999 covered the period, September 12, 1998, to May 21, 1999.
[11] Sites excluded from performance measurement included Santa Ana, Los Angeles, and Atlanta.
[12] Chicago, Honolulu, Philadelphia, and Orlando used hand-held scanners exclusively.
[13] Two fixed scanners at Boston were inoperable.
[14] Honolulu, Philadelphia, and Orlando had inefficient induction systems.

As shown below, mail was piled three to four pieces high on the induction belt at one location,[15] making it impossible to scan all applicable pieces.

The number of employees assigned to scanning operations at the ten sites was also inconsistent. For example, three people scanned mail at the site[16] with the second highest Priority Mail volume, while 15 people scanned mail at the site[17] with the third highest volume. Another facility[18] assigned an employee to scan Priority Mail at one location, although carriers unloaded mail at two locations in the

[15] Mail was stacked on induction belts at the Honolulu Airport Mail Center.
[16] Three employees were assigned to scanning operations at the Honolulu Airport Mail Center.
[17] Fifteen employees were assigned to scanning operations at the Chicago Airport Mail Center.
[18] At the Orlando Airport Mail Center, one person scanned mail, although mail was unloaded at two different locations.

facility. When multiple carriers arrived at the same time, the employee could not scan all pieces of mail. Site managers told us scanning operations were significantly impacted by the level of funding the area offices were willing to commit to performance measurement. Additionally, employees were not properly trained to use the scanning equipment.

Our review also disclosed Air Systems contract mail transported by a contractor at two sites[19] bypassed scanning operations. For example, the Northeast Area Distribution Network Office manager stated Priority Mail was transported directly to the Nashua Priority Mail Processing Center, bypassing scanning operations at the airport mail center. When mail pieces are not available to be scanned, site scan rates are distorted.

If air carriers are to be held accountable for their performance, they must possess confidence in the integrity of the Postal Service's performance measurement system. However, the six air carriers interviewed expressed a lack of confidence in the system. Specifically, they believed the scan rates were inaccurate, and therefore did not adequately represent their performance. Similarly, postal managers must also have confidence in the system if they are to commit the resources needed for performance measurement to be effective. However, managers interviewed at three[20] locations indicated that no matter how many resources they committed to scanning operations, their scan rates would not improve to the level needed for the system to be reliable.

[19] Contractor mail bypassed the Boston and Milwaukee Airport Mail Centers.
[20] Boston, Dallas, and Honolulu managers did not believe scan rates were reliable.

Accuracy of Data	Data supporting air carrier performance is inaccurate because (1) employees responsible for scanning operations did not always reset clocks to capture actual delivery times, and (2) scan rates for testing on-time delivery were inaccurate due to reliance on historical versus actual mail volumes and inclusion of air contract tagged mail. Without accurate data the Postal Service cannot hold carriers accountable for their performance or initiate corrective actions needed to improve on-time delivery.
Delivery Times	Establishing mail delivery times is critical to measuring air carrier performance. Because five[21] of the ten sites visited did not always scan incoming mail at the time of delivery, lag times of 20 minutes to 4 hours occurred between tender of the mail to the Postal Service and scanning for performance measurement. Because such delays can significantly impact the accuracy of on-time measurement, resetting scanner clocks is critical to reliable performance measurement. At four[22] of the ten sites visited, Postal Service employees did not appropriately reset clocks to reflect actual delivery times. Generally, this practice resulted in the reporting of delivery times that were one to four hours later than actual delivery. At one[23] of the five sites, postal employees changed the clocks, every hour on the hour, which consistently distorted delivery times. Facility managers indicated clocks were not changed because employees either did not know how to or did not remember to reset the clocks. Managers also believed the manual resetting of scanner clocks was inefficient. Recognizing that facilities were not always capturing actual delivery times when scanning the mail, headquarters issued a directive to performance measurement coordinators reinforcing the need to correctly set scanner clocks. Despite this directive, field sites we visited had not taken action to address this issue.

[21] Boston, Chicago, Dallas/Fort Worth, Honolulu and Philadelphia Airport Mail Centers/Facilities did not always scan incoming mail at the time of delivery.
[22] Employees at the Boston, Dallas/Forth Worth, and Honolulu and Kansas City Airport Mail Centers/Facilities did not appropriately reset scanner clocks.
[23] Employees at the Kansas City Airport Mail Center reset scanner clocks every hour on the hour.

Scan Rates

Our review disclosed performance measurement scan rates were not accurate. Because scan rates are derived from historical incoming mail volumes, to the extent average historical volumes differ from actual arrival volumes, scan rates may be misrepresented. As a result, seasonal peaks and monthly variations in mail volumes significantly distort scan rates. For example, a manager at one location[24] reported scanning mail in excess of expected arrival volumes at various times during FY 1999. Also postal managers at two locations[25] reported scan rates less than 100 percent even though every piece of mail had been scanned. As a result, mail pieces sampled are not always representative of arriving mail volumes on test days.

Further, as many as five carts of mail with air contract tags were scanned at one location,[26] although they were not counted in expected incoming mail volumes contributing to scan rates. Although air contract tags can be scanned, they are not machine-readable. The manager of the performance measurement system told us all mail should be assigned dispatch and routing tags so they can be scanned for performance measurement. The mail with air contract tags primarily came from major mailers who did not have the equipment necessary to create dispatch and routing tags.

Because performance rankings are dependent upon the volume of incoming mail scanned, it is important that scan rates be accurately reported. For example, a scan rate below 67 percent automatically places a carrier in the top performance ranking for the market pair because a true ranking cannot be determined. This can result in the shifting of mail volumes to poor performing carriers.

[24] Dallas Airport Mail Center scanned mail in excess of expected arrival volumes.
[25] The Boston and Honolulu Airport Mail Centers/Facilities scanned every peace of mail, but reported scan rates of less than 100 percent.
[26] Air contract tags at the Milwaukee Airport Mail Center could not be scanned.

Recommendations	Because the performance measurement system does not provide the type of information needed to improve air carrier delivery, performance scores are not representative of air carrier performance, and supporting data is inaccurate, we recommend the chief operating officer: 1. Suspend use of the current system.
Management's Comments	Management agreed with our recommendation and indicated the current performance measurement system will be suspended, effective February 26, 2000.
	2. Either redesign performance measurement to assess all activities impacting air carrier performance, or explore other alternatives for improving delivery performance. Management stated they would explore other options to improve air carrier performance.
Evaluation of Management's Comments	Management's comments are responsive to our findings and recommendations. Their plans to suspend the performance measurement system and explore other options should lead to a more effective alternative for improving air carrier performance.

APPENDIX A.
Air Carrier Performance Measurement
Scan Rates for the First Three Quarters of FY 1999

Area	Site	Scan Rates For FY99[27]
ALLEGHENY	Cleveland, Ohio	86%
	Columbus, Ohio	52%
	Cincinnati, Ohio	81%
	Dayton, Ohio	71%
	Harrisburg International Airport., Pennsylvania	0%
	Philadelphia, Pennsylvania	41%
	Pittsburgh, Pennsylvania	48%
CAPITAL METRO	Baltimore, Maryland	78%
	Washington National, District of Columbia	89%
	Dulles International, District of Columbia	56%
GREAT LAKES	Metro Wayne County, Michigan	53%
	Grand Rapids, Michigan	87%
	Indianapolis, Indiana	97%
	O'hare International, Illinois	72%
MIDATLANTIC	Columbia, South Carolina	75%
	Charlotte, North Carolina	86%
	Greensboro/High Point/Winston-Salem, North Carolina	71%
	Norfolk/Virginia Beach/Williamsburg, Virginia	86%
	Raleigh/Durham, North Carolina	79%
	Richmond/Williamsburg, Virginia	87%
	Louisville, Kentucky	84%
MIDWEST	Des Moines, Iowa	88%
	Wichita, Kansas	74%
	International, Missouri	95%
	Milwaukee, Wisconsin	81%
	Minneapolis/St. Paul, Minnesota	54%
	Omaha, Nebraska	94%
	St. Louis, Missouri	76%

[27] [1] Average weekly scan rates for the first three quarters of FY 1999 (accounting periods 1 through 9).

NORTHEAST	Albany, New York	0%
	Bradley International, Connecticut	55%
	Boston, Massachusetts	66%
	Buffalo, New York	78%
	Providence, Rhode Island	0%
	Portland, Maine	0%
	Rochester, New York	56%
	Syracuse, New York	81%
NY METRO	Newark International, New Jersey	36%
	Kennedy International, New York	36%
	New York La Guardia Airport, New York	59%
	San Juan, Puerto Rico	63%
PACIFIC	Honolulu, Oahu, Hawaii	69%
	Los Angeles, California	31%
	Oakland, California	77%
	Ontario, California	54%
	San Diego, California	97%
	San Francisco, California	50%
	Sacramento International, California	90%
	Orange County, California	0%
SOUTHEAST	Atlanta, Georgia	N/A[28]
	Birmingham, Alabama	75%
	Nashville, Tennessee	47%
	Ft. Lauderdale, Florida	22%
	Jackson, Mississippi	75%
	Jacksonville, Florida	68%
	Orlando International, Florida	66%
	Memphis, Tennessee	60%
	Miami, Florida	39%
	West Palm Beach, Florida	84%
	Tampa/St. Petersburg, Florida	86%
	Knoxville, Tennessee	43%
SOUTHWEST	Dallas/Fort Worth, Texas	73%
	Intercontinental, Texas	N/A
	Little Rock, Arkansas	74%
	New Orleans, Louisiana	72%
	Oklahoma City, Oklahoma	90%
	San Antonio, Texas	79%
	Tulsa, Oklahoma	83%

[28] N/A - Site data was not available for each weekly period.

WESTERN	Albuquerque, New Mexico	82%
	Anchorage, Alaska	100%
	Boise, Idaho	N/A
	Denver, Colorado	31%
	Spokane, Washington	88%
	Las Vegas, Nevada	N/A
	Portland, Oregon	73%
	Phoenix, Arizona	75%
	Reno, Nevada	78%
	Seattle/Tacoma, Washington	28%
	Salt Lake City, Utah	51%
	Tucson, Arizona	61%
	NATIONAL	65%

APPENDIX B.
Management's Comments

NATIONAL MAIL TRANSPORTATION PURCHASING

UNITED STATES
POSTAL SERVICE

January 31, 2000

RICHARD F. CHAMBERS

THRU: KEITH STRANGE

SUBJECT: Transmittal of Draft Audit Reports – Commercial Air Carrier Performance: Performance Measurement System (Report Number TR-AR-00-Draft) and Commercial Air Carrier Performance: Pay for Performance (Report Number TR-AR-00-Draft)

We are in agreement with your findings and recommendations related to the Performance Measurement System and Pay for Performance audit reports.

Based on your audit report on the Performance Measurement System, a series of issues we have been working, and the continued poor performance of the commercial air carriers, we have reconsidered the continued use of the Performance Measurement System. We will suspend the use of the present Performance Measurement System effective February 26, 2000. Based on our analysis of air transportation, we see commercial air carriers as key stakeholders in the movement of mail. Therefore, we will work with our internal customers and suppliers to explore other options to improve on-time performance and strengthen other aspects of our business.

In response to the second report on Pay for Performance, we are commencing discussions with commercial air carriers to discontinue incentive payments from the current pay-for-performance system. Although we may consider performance incentives in future contracts they will not rely on the current Pay for Performance system.

Please give my appreciation to your audit staff! We expect their findings will help us improve our efforts to meet the expectations of our customers.

J. Dwight Young, Manager
National Mail Transportation Purchasing

Attachments

cc: Mr. Black

475 L'ENFANT PLAZA SW
WASHINGTON DC 20260-6210
202-268-4377
FAX: 202-268-4313

Major Contributors to This Report