NATIONAL TRANSPORTATION SAFETY BOARD

Office of Research and Engineering

Washington, D.C. 20594

March 18, 2002

Statistical End Play Data Study

ACCIDENT

Location: Pacific Ocean near Point Mugu, California

Date: January 31, 2000

Time: 1621 Pacific Standard Time

Aircraft: MD-83, N963AS

NTSB#: DCA00MA023

INTRODUCTION

On January 31, 2000, at about 1621 PST, Alaska Airlines flight 261 a Boeing MD-83, N963AS, crashed approximately 2.69 miles north of Anacapa Island, California into the Pacific Ocean. The flight, from Puerto Vallarta, Mexico to Seattle Washington with an intermediate stop in San Francisco, was operating under title 14 CFR part 121. All 83 passengers and 5 crewmembers were fatally injured and the aircraft was destroyed. Visual meteorological conditions prevailed at the time of the accident.

Following the accident, the Federal Aviation Administration (FAA) issued a series of Airworthiness Directives (ADs) that required inspection of horizontal stabilizer trim systems for general condition and measurements of end play[1] between the jackscrew and acme nut within these systems. Initially, a telegraphic AD (2000-03-51) was issued on February 11, 2000, to all known U.S. owners and operators of Model DC-9, Model MD-90-30, Model

[1] End play is a measure of the axial distance that the jackscrew can move within the acme nut.

MD-88, and Model B-717-200 airplanes. An amendment to AD 2000-03-51, published in the *Federal Register* on February 28, 2000, made the AD effective for all operators of these aircraft beginning March 6, 2000. Another amendment, AD 2000-15-15, published in the *Federal Register* on August 8, 2000, modified the reporting procedures and clarified the procedures for inspecting and measuring end play in the jackscrew assembly.

The ADs mandated that the jackscrew assemblies be inspected for evidence of wear at regular intervals. Inspections were required prior to 650 hours total time-in-service, or within 72 hours after the effective date of the AD. The inspections were to be repeated at intervals not to exceed 650 flight hours and required the following procedures:

- Perform a general visual inspection of the jackscrew assembly for the presence of metal shavings and flakes,
- Perform a general visual inspection of the jackscrew assembly for the presence of corrosion, pitting or distress,
- Check the condition of the jackscrew assembly lubricant and, if necessary, lubricate the assembly,
- Perform a general inspection of the upper and lower mechanical stops of the jackscrew, and
- Perform a test of the horizontal stabilizer shutoff controls and, if necessary, adjust the horizontal stabilizer trim system.

The AD also mandated performance of a “wear check” within 2,000 flight hours of the most recent check, or within 30 days after receipt of the AD. The wear checks, commonly known as “end play” checks, were to be repeated at intervals not to exceed 2,000 flight hours and included the following:

- Measuring the end play of the jackscrew and acme nut, and

- Reporting results from damaged jackscrew assemblies within 48 hours and from nondamaged assemblies within 10 days.

While operators had not been required to report the results of end play measures prior to the AD, they had always been required to perform end play checks as a means to assess the degree of wear on the acme nut. The end play measure is determined by utilizing a restraining fixture, dial indicator and various brackets and clamps to measure the space between the acme nut threads and the jackscrew threads. The end play measure is typically used to infer the amount of wear that has occurred on the threads of the acme nut. Removal and replacement of the jackscrew assembly is required by the FAA when the end play measure exceeds 0.040 inch. While most operators have adopted the FAA standard, certain operators utilize a smaller criterion (e.g., .038 or .031).

Measurement Reliability

The fact that end play was measured at repeated intervals for individual jackscrew assemblies made it possible to assess whether the end play measure was reliable. Reliability refers to a measure's freedom from unsystematic errors of measurement.[2] A measure's reliability is a necessary condition for its validity – that is, if a measure is not reliable, then one can never be certain that it is measuring what it was intended to measure.

Measurement reliability is often assessed using a test-retest reliability method. That is, two consecutive measurements are recorded and, if the entity being measured has not changed (or has changed in a consistent manner) during the measurement interval, subsequent correlation of the two measures should reveal a strong relationship. The absence of such a correlation would suggest that other variables, such as measurement error, have caused variability in the observed measure.

[2] Cascio, W. F. (1991). *Applied psychology in personnel management*. Englewood Cliffs: Prentice Hall.

For the end play measures, very little change was anticipated to occur over the 2,000 flight hour interval between checks. According to The Boeing Company (Boeing) estimates, end play is expected to increase at the rate of about 0.001 inch per 1,000 flight hours. Because this expected change is so small and unidirectional, we determined that it was appropriate to utilize the test-retest method to assess measurement reliability.

In addition to the repeated measures of end play mandated by AD 2000-03-51, Boeing committed to collecting and reporting data on jackscrew assemblies that were returned to the manufacturer's[3] overhaul facility during the 2000 calendar year. At the manufacturer's only contracted overhaul facility, jackscrew assemblies are cleaned and end play checks are conducted in a controlled setting. This type of end play check is known as a "bench-check" and is believed to provide a more accurate representation of true end play compared to the "on-wing" end play check that is conducted by maintenance personnel in the field.

The first section of the report will focus on the AD-mandated on-wing end play measures that were collected between February 2000 and June 2001. We will describe the techniques that were used to screen and prepare the data for statistical analysis as well as providing an overview of the methods that were employed to assess the reliability of the end play measure. The second section of the report will be devoted to an examination and analysis of the small subset of cases where both on-wing and bench-check end play measurement data are available. At the end of each section, we will provide results and interpretations of our statistical analyses.

---

[3] Integrated Aerospace is the original equipment manufacturer and sole supplier of the jackscrew assembly. The company also has a separate overhaul facility and performs the majority of all jackscrew overhauls because of its unique ability and authorization to provide a new acme nut and screw during overhaul.

DETAILS OF THE STUDY

Section I. On-Wing End Play Data Collected by Operators

Operators were instructed to send the results of their inspections and end play checks – as well as aircraft fuselage numbers, flight hours, and flight cycles accumulated – to Boeing, who consolidated all files that were submitted electronically. The electronic database, as well as any records that were submitted in paper form, were sent to the Safety Board. The Safety Board received data sets from Boeing on three occasions between March and June 2001.

Preliminary Data Screening and Analysis

In May 2001, the Safety Board screened and analyzed the data that had been received to date. At that time, the Safety Board had received a total of 1,970 cases from 1,244 aircraft representing 35 operators. During the screening process, several errors were discovered including incorrect dates, duplicate records, and multiple measurement conventions (i.e., metric and imperial). The Safety Board described these problems and presented preliminary results at a Systems Group meeting held on May 31, 2001. The Group concluded that, due to changes in the end play measurement process that had occurred since the original AD, additional data should be gathered from the operators before conducting a final analysis.

Final Data Screening and Analysis

On June 11, 2001, the Safety Board contacted Boeing to request all additional electronic data submissions that had been received since the previous reporting date. Boeing responded by sending a compact disc containing an electronic database of all information they had gathered to date as well as several envelopes containing hard-copy data. Boeing representatives remarked that, in addition to adding new records, they had also modified several records due to previous processing errors. The electronic data set received from

Boeing in June contained 2,771 records. After adding 403 hard-copy records, there were a total of 3,174 records from 1,493 aircraft representing 44 operators.

As noted in the introduction, the AD required repetition of the end play checks at intervals not to exceed 2,000 flight hours. This resulted in multiple records for many aircraft. Table 1 provides an overview of the number of aircraft, for both the preliminary and final analyses, for which multiple end play measures were received. For a detailed account of the final number of cases that were received from each operator, including the percentage of the fleet represented in the sample, see appendix A.

Table 1. The number of aircraft measured on each occasion for both preliminary and final data analyses.

Measurement Interval	**Preliminary**	**Final**
Time 1	1,244	1,493
Repeated at Time 2	510	931
Repeated at Time 3	163	539
Repeated at Time 4	40	129
Repeated at Time 5	10	66
Repeated at Time 6	3	13
Repeated at Time 7	0	3
Total	1,970	3,174

The core variables utilized in this study were: Aircraft fuselage number, which served as a unique identifier for each aircraft; total number of flight hours at time of the end play check, which allowed us to track the duration of the intervals between measurements; and the end play measure itself.

Prior to both the preliminary and final analyses, data were screened for several types of errors. The following sections provide an overview of the screening techniques that were used and an explanation of how they affected the resulting data set.

Missing Data

When one of the core variables was missing for a given record, efforts were made to obtain the correct information. For example, using the Airclaims[4] database, aircraft serial numbers were used to determine corresponding fuselage numbers. When flight hours were missing, Boeing provided an estimate based on operator-reported monthly flight hours corresponding to the end of the calendar month in which the end play check took place. For example, if a check occurred on May 18, the flight hours recorded on May 31 were used to replace the missing field.

Duplicate Cases

In the preliminary data set, there were 133 cases where all of the core variables were identical except for the date of measurement. Since this is virtually impossible, these cases were removed from the data set before the preliminary analysis. In the final data set, there were no duplicate cases since Boeing had detected and removed them prior to sending the data set to the Safety Board.

Date of Measurement

In the preliminary data set, there were 215 cases from one operator in which the date of the end play check was reported as occurring during the years 1996 and 1997, several years prior to the issuance of the AD. Close scrutiny of these data revealed that there had been a data processing problem with the spreadsheet program, Microsoft Excel, which had erroneously modified the originally recorded dates by exactly 4 years and 1 day. Therefore, for the final set, these data were corrected to reflect their actual dates.

Before sharing data with the Safety Board, Boeing also removed cases where reported dates occurred after the report was submitted (i.e., in the future). They also removed cases

[4] Airclaims Limited: Client Aviation Enquiry Database, Version 2.0 (CASE2).

where both date and flight hours fields were missing. Finally, when necessary, both Boeing and the Safety Board converted dates to a conventional month/date/year format.

Of the 3,174 cases in the final sample, there were 307 cases with no date information. However, since date is not a core variable, these cases were not removed. In addition, there were 15 cases in which the recorded date suggests that the measurement took place *before* the original AD was issued on February 11, 2000. In 9 of these cases, the measure was conducted 1 day before the AD and none of the remaining 6 cases took place more than 3 months prior to the AD. Because these cases occurred within a relatively brief time period of the accident, they were not removed from the data set.

In many cases, the initial end play measure reported for a given aircraft took place after the 30-day window of time required by AD 2000-03-51. For example, there are 282 cases with dates after September 6, 2000, (6 months after the effective date of the AD) that represent the initial or the only entry for a given aircraft. However, apart from being belated, there was no reason to believe that these measures were erroneous. Therefore, they were not removed from the data set.

Multiple Measurement Conventions

While most operators reported their end play measurements in inches (i.e., an imperial unit), several operators reported end play in millimeters (i.e., a metric unit). In some cases, Boeing converted metric units to imperial, and in some cases the conversions were done at the Safety Board. Measurements reported in millimeters were converted to inches using 0.03937 as a multiplying factor.

Out-of-Specification End Play Measures

In the preliminary data set, there were 98 cases in which end play measures of less than 0.003 inch were documented. Because the manufacturer-specified initial tolerance for the end play measure is between 0.003 and 0.010 inch, it is unlikely that measures less than 0.003 inch were representative of true end play in the assembly. After the data from the preliminary phase had been analyzed, Boeing discovered that, in the case of one operator, there were 43 cases below 0.003 inch where the end play measure had been transposed with another measure, free play,[5] which is generally much smaller than end play.

Boeing checked for problems with data transposition and made corrections before submitting the final data set to the Safety Board. Nevertheless, in the final set, there remained 71 cases in which end play measures of less than 0.003 inch were reported. None of these cases were removed from the final data set.

Decreasing Flight Hours

In the preliminary data set, there were 72 cases in which the flight hours recorded on a given date were greater than the flight hours recorded on a subsequent date. Since it is not possible for flight hours to decrease, it is likely that this information was recorded in error.

During the interim between the analysis of the preliminary data set and the delivery of the final data set, Boeing detected and corrected certain errors that reduced the prevalence of flight hours that decreased over time. For instance, in several cases, the flight hours for an individual jackscrew had been recorded instead of the flight hours for the fuselage. In other cases, the date of measure was entered incorrectly. In spite of these corrections, the final set contained 26 cases where flight hours decreased between one time and the next. None of these cases were modified or deleted prior to the final analysis.

[5] Free play is a measure of the amount of movement in the torque tube drive bearing contained within the gearbox support.

Incorrect Fuselage Numbers

In the preliminary data set, the Safety Board detected two cases where the incorrect fuselage number was used to identify an aircraft. In one case, the fuselage number reported did not match any known aircraft in the Airclaims database. In the other case, one operator reported a fuselage number that, according to Airclaims, actually belonged to another operator. Because the true ownership of the fuselage was impossible to determine in these cases, they were removed from the preliminary data set prior to analyses. Finally, before submitting the final data set, Boeing removed two additional cases with incorrect fuselage numbers.

Changed Jackscrews

The physical mechanisms that underlie the operation of the jackscrew assembly suggest that, with time and use, the acme nut will wear causing an increase in measured end play. Boeing has estimated that end play will increase at a rate of approximately 0.001 inch per 1,000 hours flown by the aircraft. Therefore, since operators reported end play measures at intervals of approximately 2,000 flight hours (as mandated by the AD), we expected to observe increases of approximately 0.002 inch per interval.

If, over the course of our investigation, a jackscrew assembly were replaced without our knowledge, it would pose a problem for the analysis. Contrary to our prediction of a gradually increasing end play measure, a changed jackscrew would introduce a substantial reduction in end play between one measurement and the next. To avoid this problem, it was important to screen out any jackscrew assemblies that had been replaced during the data collection period.

In the data received from Boeing, operators provided information about jackscrew assemblies that had been removed and/or replaced. In the final set, there were 48 cases in which a jackscrew change had been documented. These cases were removed from the final analysis. In addition, we utilized a mathematical convention to estimate whether a jackscrew assembly had been changed during the course of the study. The maximum end play threshold used by all operators in the United States for jackscrew replacement is 0.040 inch, which is the standard stated by the FAA. However, certain operators use a more stringent criterion (for example, one operator uses a threshold of 0.031 inch). A jackscrew with a measure exceeding the operator's criterion for replacement, followed by a measure in which the end play fell between 0.003 and 0.010 inch, was considered to have been replaced. Using this rule, we identified 58 additional cases for a total of 106 cases that were removed from the final analysis set.

Description of Reliability Analysis

After screening the data for errors, the resulting set contained 1,388 cases at Time 1, 852 cases at Time 2, and 482 cases at Time 3. To assess whether the end play measure was reliable, we used a test-retest method. End play measures reported on three consecutive occasions for the same aircraft were compared using a correlation analysis which measures the relationship between two sets of numbers and produces a value known as the "correlation coefficient" (represented in notation as "r"). The correlation coefficient can range between ±1.0. A coefficient of +1.0, known as a perfect positive correlation, means that in each case changes in one measure resulted in an identical change in the second measure. A coefficient of -1.0, known as a perfect negative correlation, means that changes in one measure resulted in an identical change in the other measure, but that the change was in the opposite direction. A coefficient of zero means there was no relationship between the two measures and that a change in one had no effect on the other.

In the current analysis, we predicted that the resulting correlation coefficient would be close to +1.0 suggesting a high degree of relationship between end play measures taken on consecutive occasions. For example, if the measured change in end play over the course of 2,000 flight hours ranged from no change to a change of 0.002

inch, then the result would be a correlation coefficient of approximately +0.992.[6] The squared value of the correlation coefficient, known as the "coefficient of determination" and represented in notation as "R^2", indicates the amount of shared variability between the two sets of measures. In the case of r= +0.992, R^2 would be 0.984 indicating that 98 percent of the variability in one measure can be predicted by variability in the other measure.

Results of Reliability Analysis

The results from the correlation analysis (after data cleaning and removal of changed jackscrews) for three consecutive sets of end play measurements are presented in table 2. The correlation coefficient for the first and second measurements was +0.553 and for the second and third measurements was +0.416. The coefficients of determination, representing shared variance between sets of measures, were 0.306 and 0.173 respectively, suggesting that measurement reliability is low.

Scatter plots representing the relationships between consecutive measurements are depicted in figures 1 and 2. If there were no change in measured end play between two consecutive occasions, the result would be expressed as a 45-degree diagonal line and the correlation coefficient would be +1.0. However, the actual points shown in figures 1 and 2 reflect a broader dispersion consistent with the lower correlation coefficients that were produced by the analyses.

[6] A normally distributed set of 100 cases was randomly generated using a mean and standard deviation representative of the actual set of end play measures. To each number, we added another randomly generated number with mean of 0.002 and standard deviation of 0.001. The resulting correlation of these two sets of numbers was +0.992.

Table 2. Correlation coefficients (r) between end play measures taken on three consecutive occasions.

	Measure 1	Measure 2	Measure 3
Measure 1	1.00	0.553	0.362
Measure 2		1.00	0.416
Measure 3			1.00

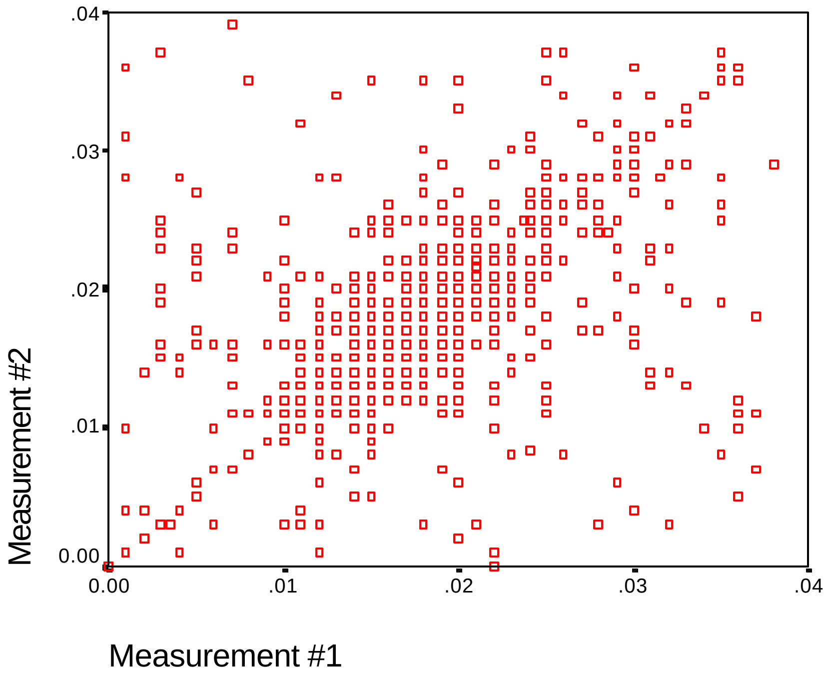

Figure 1. A scatter plot depicting the first (x-axis) and second (y-axis) sets of end play measurements (r = +0.553).

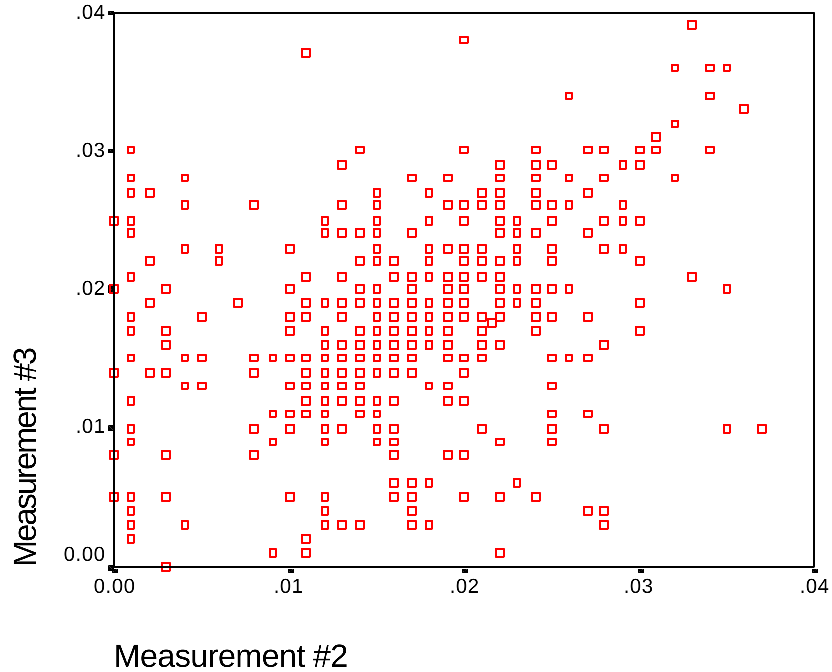

Figure 2. A scatter plot depicting the second (x-axis) and third (y-axis) sets of end play measurements (r = +0.416).

Points that appear in the upper left and lower right hand portions of figures 1 and 2 represent substantial changes in the recorded end play measures over the course of a measurement interval. Figures 3 and 4 display histograms that show the distribution of changes that took place between consecutive measurements. While in most cases the change is small, there are a number of cases where the documented change was greater than 0.020 inch, which is 10 times Boeing's estimated change during a 2,000-flight hour interval. Furthermore, there were several instances where the end play measure appeared to *decrease* over time. In theory, a decrease in end play over time is physically impossible except in cases where debris has become lodged between the threads of the jackscrew and those of the acme nut. Therefore, there is reason to conclude that other factors, besides actual change in distance between the jackscrew and acme nut threads, contributed to the observed differences in end play measures.

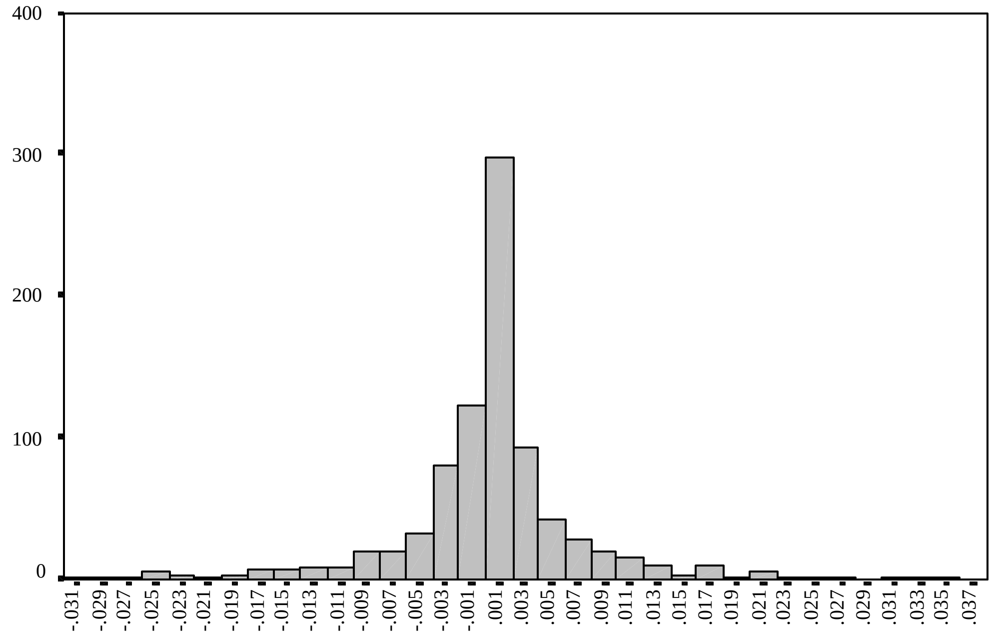

Figure 3. Histogram displaying the changes in end play between the first and second measurements.

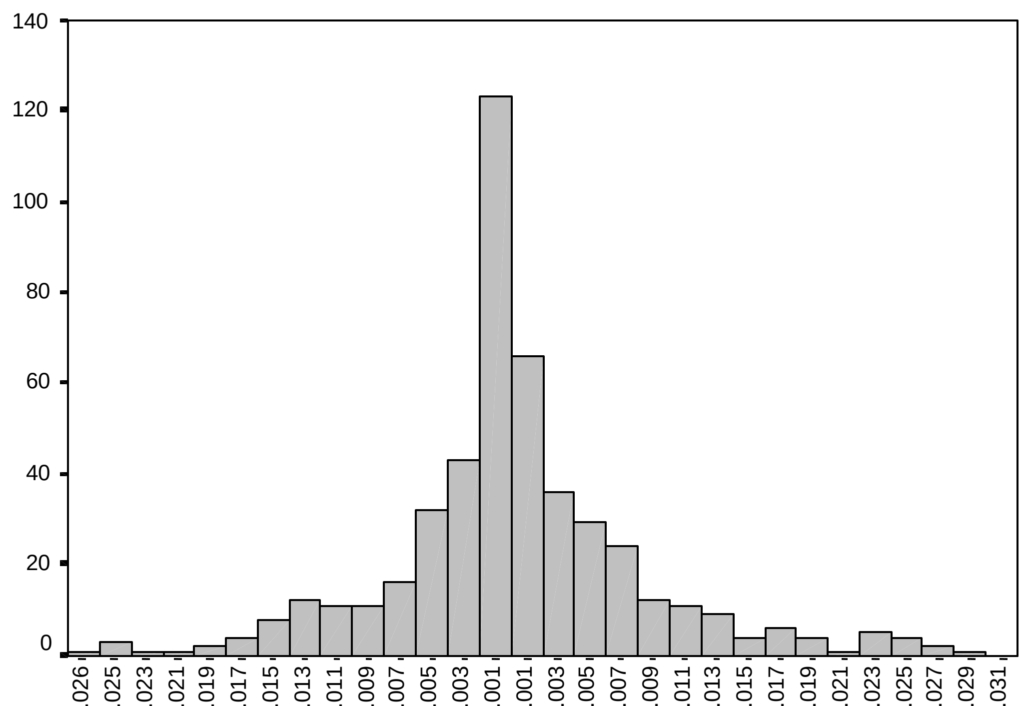

Figure 4. Histogram displaying the changes in end play between the second and third measurements.

One potential factor that may have contributed to the low test-retest correlation coefficients was the fact that, while the AD stated that the measurement interval should not exceed 2,000 flight hours, it did not mandate that the measures be taken at exactly 2,000-hour intervals. In fact, the intervals between measurements varied widely with several cases exceeding 2,000 flight hours and a few cases where flight hours actually decreased (see Section I: On-Wing End Play Data Collected by Operators.)

If the test-retest method were performed under ideal conditions, each measure would be taken at identical intervals. Because this was impossible under the current circumstances, two alternative analyses were performed. In the first, cases in which the interval between two measures was either greater than 2,000 flight hours or less than 1 flight hour were removed from the analysis and the correlation was repeated. This led to the removal of 146 cases from the analysis that compared the first and second end play measures: The resulting correlation coefficient was +0.642 (R^2 = 0.412). For the analysis comparing the second and third end play measures, 27 cases were removed from the analysis and the resulting correlation coefficient was +0.398 (R^2 = 0.158).

The second alternative[7] involved calculating the amount of change in end play that had occurred over the time interval and then normalizing or standardizing the measures based on the size of the elapsed interval. Boeing's estimated rate of wear on the acme nut is 0.001 inch per 1,000 flight hours. So for example, if the interval for a given aircraft were 1,800 flight hours, we would adjust the end play value of the second measure by subtracting 0.0018 inch. Using this logic, we normalized all end play measures in an attempt to diminish the effects of time on end play. The resulting correlation coefficient between the first and

[7] A third alternative, similar to the second, involved normalizing end play measures using a wear rate derived from actual data. This method resulted in correlations that were lower than the unadjusted correlations. In addition, using observed data to generate a wear rate implicitly assumes that the observed data is valid. This creates a circular logic and raises doubts about the utility of the analysis. Therefore these results are not included in the discussion.

second measures was +0.479 (R^2 = 0.229) and the correlation between the second and third measures was +0.393 (R^2 = 0.154). Table 3 presents the correlation results from the original analysis as well as the results using the two alternative methods.

Table 3. Correlation coefficients for the three correlation analyses performed to test the reliability of the end play measurement.

	Original analysis	Flight hour intervals of <1 and >2000 deleted	Adjusted using Boeing's wear rate
r_{1-2}	0.553	0.642	0.479
r_{2-3}	0.416	0.398	0.393

Section II. End Play Data for Jackscrews Removed from Service

Description of Data and Data Screening

Integrated Aerospace reported that 157 jackscrew assemblies were returned for overhaul during the 2000 calendar year. As shown in Table 4, this number is markedly higher than the years prior to 2000. It is possible that this increase can be attributed to the postaccident AD that required end play checks for the entire fleet of Model DC-9, MD-90-30, MD-88, and B-717-200 aircraft. While 49 of the 157 jackscrew assemblies were removed because of end play measures that exceeded operators' criteria for removal, the other reasons for removal included the presence of metal flakes, excessive free play, or damage to the assembly.

Table 4. Number of jackscrew assemblies removed per year for the years 1994-2000.

Year	Number of Units Overhauled
1994	23
1995	40
1996	47
1997	60
1998	75
1999	80
2000	157

When jackscrew assemblies are returned to an overhaul shop, they are cleaned to remove any grease or debris and the end play measure is assessed using the bench-check. A jackscrew assembly with a bench-check end play that falls within the FAA tolerance for use on an aircraft (and displays no other apparent damage) may be overhauled and returned to service. However, certain overhaul shops discard all acme nuts that are removed from service by operators.

Of the 157 jackscrew assemblies that were removed in 2000, the bench-checks for 142 units resulted in end play measures that fell within ascribed end play tolerances (between 0.003 and 0.040 inch). Twelve assemblies had bench-check end play measures greater than 0.040 inch and there were no bench-check readings reported for the remaining three.

For each jackscrew assembly that was returned to Integrated Aerospace for overhaul in 2000, Boeing contacted the operator to obtain the corresponding on-wing end play measure documented prior to its removal. They were able to obtain a matched sample (i.e., both on-wing and bench-check end play measures) for a total of 64 jackscrew assemblies; however, they were unable to obtain matching on-wing checks for 9 of the 12 units that exhibited end play measures greater than 0.040 inch during the bench-check.

Within the sample of on-wing end play measures submitted by Boeing, there were a few cases in which the documented end play measure was extremely high, suggesting the possibility of a measurement or a documentation error. In one case, an on-wing end play measure of 0.410 inch was recorded – an impossibly high number given that the tolerance ranges from 0.003 to 0.040 inch. Boeing contacted the operator and determined that a decimal error had occurred during the documentation of this record: The actual measure was 0.041 inch. Because this particular case concerned a documentation error, rather than an operator error, it was corrected prior to our analysis.

In a similar case, an on-wing end play of 0.390 inch was documented. Although it is likely that this represents a similar documentation error, Boeing was unable to confirm this information with the operator. Therefore, initially it was not removed from the analysis set. A third case involved a measurement error made by an inspecting mechanic. According to Boeing, the mechanic incorrectly preloaded the dial indicator to 0.100 inch. Upon loading the restraining fixture, the indicator read 0.076 inch. Reverse calculation notes that the actual on-wing measurement was 0.024 inch. However, because errors made by mechanics during the end play check can greatly influence decisions regarding the health of a jackscrew assembly, we did not remove this case from the analysis or modify it in any way.

After correcting the verified documentation error, we calculated the difference in end play measures between the on-wing and the bench-check settings. In 45 of 64 cases (70.3 percent) the on-wing check was greater than the bench-check. In six cases (9.4 percent) the two numbers were equal and in 13 cases (20.3 percent), the on-wing check was less than the bench-check. Table 5 provides a more detailed breakdown of the distribution of difference scores. In addition, an assessment of the on-wing end play checks that occurred *prior* to removal of the jackscrew assembly (using the larger set of reliability data) revealed several cases where sequential measurements of on-wing end play produced highly variable readings. Appendix B includes examples of these cases.

Table 5. A breakdown of the differences between the end play measure taken in the field (on-wing) and those taken at Integrated Aerospace (bench-check).

Number of Cases	**Difference Score**	**Direction of Difference**
3	+0.031 or greater	On-wing end play greater than bench-check end play
3	+0.021 to +0.030	
6	+0.011 to +0.020	
13	+0.006 to +0.010	
20	+0.001 to +0.005	
6	0	No difference
5	-0.001 to -0.005	On-wing end play less than bench-check end play
2	-0.006 to -0.010	
3	-0.011 to -0.020	
2	-0.021 to -0.030	
1	-0.031 or less	

Description of Validity Analysis

In Section I of this report we described a test-retest technique used to assess measurement reliability. Using that method, consecutive on-wing end play measures from the same jackscrew were correlated. Because we expected the change in on-wing end play over the course of 2,000 flight hours to be very small and unidirectional, we predicted a correlation coefficient near +1.0, suggesting that a change in one measure would result in a similar change in the corresponding measure.

To assess the validity of the on-wing end play measure we used a similar technique. However, instead of correlating consecutive measures of on-wing end play, we calculated the correlation between the on-wing measures and the bench-check measures. This method assumes that the bench-check measure is a “criterion” or a standard that represents the true state of the jackscrew end play. Therefore, if the on-wing end play measure is valid – i.e., represents the actual state of the end play – then the correlation coefficient representing the relationship between the on-wing and bench-check measures should be close to +1.0.

Results of Validity Analysis

The correlation analysis comparing 64 matched sets of bench and on-wing end play data resulted in a correlation coefficient of +0.172, suggesting that measurement validity is very low. A scatter plot representing the relationship between the bench-check and on-wing end play measures is represented in figure 5. It is clear from this figure that the on-wing measure of 0.390 is an outlier and is likely influencing the magnitude of the correlation coefficient. Therefore, a second correlation analysis was conducted after removing that outlier from the sample. The resulting correlation coefficient for the remaining 63 matched sets was +0.442. A scatter plot depicting the relationship between the two sets of measures, after removing the outlier, is presented in figure 6. While the removal of the outlier produces a substantial improvement in the correlation, the coefficient of determination (R^2) is 0.195, suggesting that measurement validity is much lower than expected.

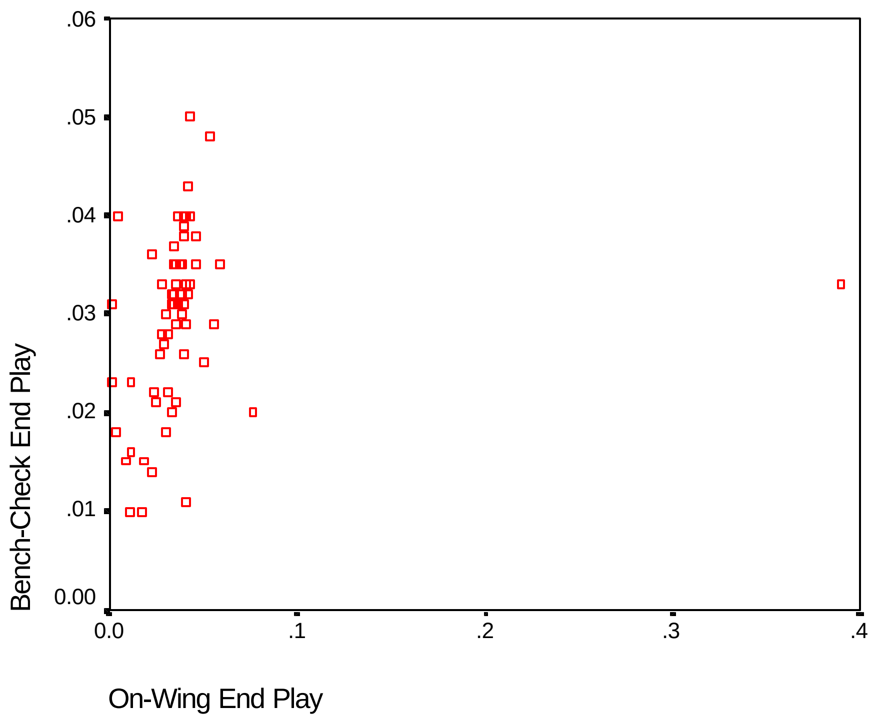

Figure 5. A scatter plot depicting the on-wing (x-axis) and bench-check (y-axis) for the 64 matched sets of end play measures (r = +0.172).

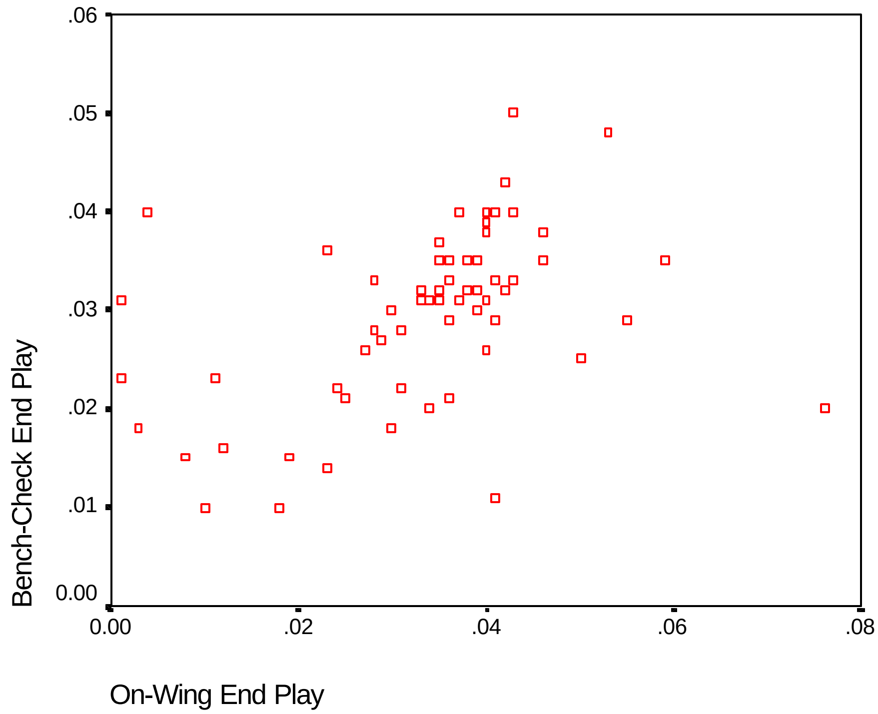

Figure 6. A scatter plot representing 63 on-wing (x-axis) and bench-check (y-axis) measures after removing one outlier measure (r = +0.442).

CONCLUSION

The goal of this study was to assess the reliability and validity of the on-wing end play measure. Reliability refers to a measure's repeatability or its freedom from measurement error, and it is a necessary condition for validity. Specifically, if a measure is unreliable, one can never be certain that it is measuring what it was intended to measure.

In the case of the on-wing end play measure, the test-retest reliability method showed that there were relationships between consecutive measures, yet the correlations were very low by the standards of measurement reliability. Two alternative analyses, designed to reduce the effects of widely varying time intervals between measures, produced mixed results. In the analysis that removed cases with extreme flight hour intervals, (i.e., greater than 2,000 or less than 1) the correlation coefficient representing the first and second sets of measurements (+0.642) increased compared to the original analysis (+0.553). However, the corresponding correlation coefficient for the second and third sets of measurements (+0.398) decreased compared to the original (+0.416). The second alternative, using Boeing's estimated wear rate to standardize measures based on the elapsed intervals between on-wing checks, resulted in correlation coefficients that were lower in both cases.

In summary, all three test-retest correlation analyses conducted for this study suggest that there is a very large amount of measurement error present in the end play measure. Using the "best case" correlation coefficient of +0.642, the amount of shared variability between two consecutive measures, or R^2, indicates that only 41 percent of the variability is explained.

The second section of this study focused on a set of 64 matched on-wing and bench-check end play measures for jackscrew assemblies that were removed for a variety of reasons. The end play measure obtained from the bench-check, which is conducted in a controlled laboratory environment, represents our best estimate of true end play. Therefore, even though the set of matched cases was relatively small, it allowed for a more direct analysis of the validity of the on-wing end play check.

In more than 70 percent of the end play measures, the on-wing check was greater than the bench-check with differences ranging from 0.001 to 0.357 inch.[8] Boeing representatives have suggested that this shows a "conservative trend in the procedure." In other words, they contend that the on-wing check is designed to err in a direction that will lead to more "false alarms" (i.e., jackscrews that are tested above 0.040 but are actually at or below 0.040) and fewer "misses" (i.e., jackscrews that are tested at or below 0.040 but are actually above 0.040).

[8] The difference of 0.357 is attributable to an outlier where the on-wing check was recorded as 0.390 inch. The next lowest difference was 0.056.

In theory, a test designed to produce more conservative (i.e., larger) end play measures in a consistent fashion might serve to increase the probability of false alarms and decrease the probability of misses. However, Boeing representatives have stated that this was not their intent when designing the on-wing check. Because the two tests follow the same basic procedure, it is reasonable to expect that the tests conducted in both settings would be approximately equal most of the time. In fact, observed differences between the bench-check and on-wing measures are sometimes quite extreme. In six cases (9.4 percent) the on-wing measure is greater than the bench-check measure by 0.021 inch or more and, more importantly, in three cases (4.7 percent) the wing measure is less than the bench measure by 0.021 inch or more.

While many possible factors could contribute to the differences in these measures, two likely causes are presented here. First, for the on-wing test, a significant amount of force (approximately 2,000 pounds) is applied to the jackscrew during the measurement procedure in order to offset the weight of the horizontal stabilizer and to compress the grease that may be trapped between the screw and nut threads. This force may cause a slight deflection of the threads and increase the on-wing measure by a few thousandths of an inch. The fact that a similar force is not applied in the bench-check setting may contribute to the higher end play measures witnessed in the on-wing check compared to the bench-check.

However, the second factor that may lead the two measures to be different affects the difference in the opposite direction. When jackscrew assemblies are measured on the wing, there is a chance that grease and or other debris may impede the movement of the screw relative to the nut, thus causing a smaller end play measurement. However, when a jackscrew is removed, the overhaul shop thoroughly cleans the assembly before measuring end play in the bench setting. Thus, the effects of the cleaning may increase the end play measured on the bench as compared to that measured on the wing.

In conclusion, the on-wing end play check is currently used to establish whether or not a jackscrew assembly should be removed from an aircraft. This study focused on assessing the reliability and validity of the on-wing check. In the end, both the reliability and validity analyses conducted for this study suggested that there is a large amount of measurement error present in the on-wing end play check. In the absence of additional information such as the rate of acme nut wear and the thread thickness at which failure may occur, the observed level of measurement error raises doubts about the utility of the existing end play measurement procedure.

Jana M. Price, Ph.D.
Transportation Research Analyst

Appendix A

Number of cases received for each operator and number of repeated measures (after removing changed jackscrews).

Operator	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Total
1	268	230	140	13	0	0	0	651
2	28	2	1	0	0	0	0	31
3	1	0	0	0	0	0	0	1
4	17	5	0	0	0	0	0	22
5	2	0	0	0	0	0	0	2
6	6	6	6	5	4	4	2	33
7	81	0	0	0	0	0	0	81
8	2	0	0	0	0	0	0	2
9	46	20	0	0	0	0	0	66
10	6	1	0	0	0	0	0	7
11	28	14	4	0	0	0	0	46
12	31	7	1	0	0	0	0	39
13	12	12	10	2	0	0	0	36
14	14	0	0	0	0	0	0	14
15	59	48	13	0	0	0	0	120
16	12	10	0	0	0	0	0	22
17	12	0	0	0	0	0	0	12
18	132	114	62	4	0	0	0	312
19	5	5	1	0	0	0	0	11
20	1	0	0	0	0	0	0	1
21	12	2	0	0	0	0	0	14
22	45	14	0	0	0	0	0	59
23	5	0	0	0	0	0	0	5
24	41	40	19	3	0	0	0	103
25	2	0	0	0	0	0	0	2
26	1	0	0	0	0	0	0	1
27	1	0	0	0	0	0	0	1
28	34	28	1	0	0	0	0	63
29	24	3	0	0	0	0	0	27
30	132	130	129	73	49	7	1	521
31	6	6	6	1	0	0	0	19
32	5	0	0	0	0	0	0	5
33	1	0	0	0	0	0	0	1
34	81	1	0	0	0	0	0	82
35	1	0	0	0	0	0	0	1
36	19	0	0	0	0	0	0	19
37	18	4	0	0	0	0	0	22
38	117	108	74	1	0	0	0	300
39	11	0	0	0	0	0	0	11
40	64	43	15	5	0	0	0	127
41	2	0	0	0	0	0	0	2
42	3	0	0	0	0	0	0	3

Operator	Time 1	Fleet	% Fleet
1	268	284	94%
2	28	74	38%
3	1	3	33%
4	17	18	94%
5	2	5	40%
6	6	13	46%
7	81	89	91%
8	2	3	67%
9	46	55	84%
10	6	6	100%
11	28	33	85%
12	31	35	89%
13	12	12	100%
14	14	13	108%
15	59	72	82%
16	12	14	86%
17	12	12	100%
18	132	136	97%
19	5	7	71%
20	1	10	10%
21	12	15	80%
22	45	62	73%
23	5	29	17%
24	41	41	100%
25	2	32	6%
26	1	1	100%
27	1	n/a	n/a
28	34	34	100%
29	24	27	89%
30	132	180	73%
31	6	5	120%
32	5	8	63%
33	1	3	33%
34	81	105	77%
35	1	1	100%
36	19	32	59%
37	18	26	69%
38	117	142	82%
39	11	13	85%
40	64	77	83%
41	2	9	22%
42	3	n/a	n/a

Appendix B

For each aircraft documented in the set of data that contained both on-wing and bench-check data, we utilized the fuselage number to track the history of end play checks that occurred prior to the removal of the jackscrew assembly (using the larger set of reliability data). This search revealed several cases where sequential measurements of on-wing end play produced highly variable readings. Examples of these cases are listed below.

Case #1

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	2/17/00	0.030	
On-Wing	2/28/00	0.040	"Wore" 0.010 in 11 days
Bench	3/7/00	0.040	

Case #2

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	2/10/00	0.001	
On-Wing	3/6/00	0.001	
On-Wing	6/2/00	0.024	"Wore" 0.023 in 3 months
On-Wing	8/10/00	0.024	
On-Wing	None	0.040	Failed end play check
Bench	9/29/00	0.031	

Case #3

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	2/11/00	0.001	
On-Wing	2/19/00	0.001	
On-Wing	5/17/00	0.038	"Wore" 0.037 in 3 months
On-Wing	None	0.041	Failed end play check
Bench	10/18/00	0.033	

Case #4

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	None	0.039	Failed due to "metal flakes"
Bench	2/15/00	0.030	

Bench	2/16/00	0.023	Bench test changed by 0.007 after 1 day

Case #5

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	2/11/00	0.025	
On-Wing	2/27/00	0.025	
On-Wing	9/15/00	0.009	"Grew" 0.016 in 7 months
On-Wing	None	None	Removed for "nicks and gouges"
Bench	10/6/00	0.020	

Case #6

Type of End Play Check	Date	End Play Measure	Notes
On-Wing	2/14/00	0.036	
On-Wing	7/14/00	0.012	"Grew" 0.014 in 5 months
On-Wing	None	0.036	"Worn pin out of tolerance"
Bench	12/5/00	0.021	