BEFORE THE

DEPARTMENT OF THE TREASURY

Office of Comptroller of the Currency
Docket No. 06-04; RIN 1557-AC89

Office of Thrift Supervision
No. 2006-06; RIN 1550-AC01

FEDERAL RESERVE SYSTEM
Docket No. R-1250

FEDERAL DEPOSIT INSURANCE CORPORATION
RIN 3064-AC99

NATIONAL CREDIT UNION ADMINISTRATION

FEDERAL TRADE COMMISSION
PROJECT NO. R611017; RIN 3084-AA94

INTERAGENCY ADVANCE NOTICE OF PROPOSED RULEMAKING: PROCEDURES TO ENHANCE THE ACCURACY AND INTEGRITY OF INFORMATION FURNISHED TO CONSUMER REPORTING AGENCIES UNDER SECTION 312 OF THE FAIR AND ACCURATE CREDIT TRANSACTIONS ACT.

COMMENTS SUBMITTED BY
THE OWNER-OPERATOR INDEPENDENT DRIVERS ASSOCIATION, INC.

JAMES J. JOHNSTON
President
Owner-Operator Independent
Drivers Association, Inc.

PAUL D. CULLEN, JR.
The Cullen Law Firm, PLLC
1101 - 30th Street, N.W.
Suite #300
Washington, DC 20007
Tel. (202) 944-8600

May 22, 2006

I. INTRODUCTION

A. PROCEDURAL STATEMENT

These comments are submitted by the Owner-Operator Independent Drivers Association, Inc. ("OOIDA" or "Association") in response to an Advanced Notice of Proposed Rulemaking and a Request for Comments pertaining to Procedures to Enhance the Accuracy and Integrity of Information Furnished to Consumer Reporting Agencies Under Section 312 of the Fair and Accurate Credit Transactions Act. The multi-agency notice of this request was published in the Federal Register on March 22, 2006.

B. THE INTEREST OF THE OWNER-OPERATOR INDEPENDENT DRIVERS ASSOCIATION, INC.

OOIDA is a not-for-profit corporation incorporated in 1973 under the laws of the State of Missouri, with its principal place of business in Grain Valley, Missouri. OOIDA is the largest international trade association representing the interests of independent owner-operators and professional drivers on all issues that affect small business truckers. The more than 140,000 members of OOIDA are small-business men and women located in all 50 states and Canada who collectively own and operate more than 210,000 individual heavy-duty trucks. Many of OOIDA's members have their own DOT authority to operate in interstate commerce. The mailing address of the Association is:

Owner-Operator Independent Drivers Association, Inc.
P.O. Box 1000
1 NW OOIDA Drive
Grain Valley, Missouri 64029
www.ooida.com

The Association actively promotes the views of small business truckers and professional drivers through its interaction with state and federal government agencies, legislatures, the

courts, other trade associations, and private businesses to advance an equitable and safe environment for commercial drivers. OOIDA is active in all aspects of highway safety and transportation policy, and represents the positions of small business truckers in numerous committees and various forums on the local, state, national, and international levels. Many consumer reporting agencies trade in credit reports containing the employment histories of commercial motor vehicle drivers. The accuracy of consumer reports containing the employment history of a truck driver can have a substantial impact on the employment opportunities available to truck drivers, including members of OOIDA, and to the pool of prospective drivers available to motor carriers.

II. SUMMARY

The Fair Credit Reporting Act ("FCRA") requires that consumer reporting agencies "follow reasonable procedures to maximize the possible accuracy of consumer reports."[1] Accuracy is a concept that implies both truth and precision. "On its maiden voyage, the Titanic sailed close to an iceberg." This statement is true, but it is not precise. Therefore, it is not "accurate." "Maximizing possible accuracy" under the FCRA, therefore, is frequently an exercise of improving the precision of statements in a consumer report, especially employment histories. Often consumer reporting agencies provide the format used by previous employers and businesses to contribute the statements that populate the consumer reports purchased by prospective employers. This rulemaking, therefore necessarily implicates how the process and format for collecting statements affects the accuracy of consumer reports.

Truck drivers are frequently the subject of consumer reports containing their employment

[1] 15 U.S.C. § 1681e(b)

histories at various motor carriers. Some statements regularly used in such reports are so imprecise that they, simultaneously, have both derogatory and neutral meanings. An example of a form used by former employers to submit statements about drivers and its guide is attached as Exhibit 1. For example, the term "cargo loss" on a driver's report can mean either that cargo was lost due to the negligence or theft of the driver or that the cargo was stolen by others or subject to weather damage through no fault of the driver. Because all of these meanings are implied, the statement is inherently imprecise. Without precision, a consumer report containing that statement alone is far from accurate. Consumer reports are frequently used by motor carriers in the defense of negligent hiring actions. Motor carrier employers, therefore, refuse to hire a driver based solely on a consumer report that contains a derogatory statement whether its author meant it to have that meaning or not.

Consumer reporting systems with such high levels of inaccuracy fail to promote the goals of the Fair Credit Reporting Act to facilitate the efficient transfer of information and promote commerce. Good truck drivers with derogatory consumer reports are effectively blackballed from the industry. Prospective employers are discouraged from hiring a good driver because an employment history connotes a bad employee. Perhaps most importantly, truck drivers believe that companies make potentially derogatory statements in their employment histories as a way of retaliating against or disciplining them for engaging in such activities as demanding to be paid on-time, refusing to drive a defective truck, or refusing to drive more than the allotted time allowed under federal hours-of-service rules.

The claims that consumer reports are used to retaliate against drivers are extremely difficult to prove and prosecute. Individual drivers have few resources to bring such actions.

OOIDA believes that such uses would be greatly curtailed, however, if persons who provide employment history information to consumer reporting agencies were required to follow appropriate standards for "maximizing possible accuracy." Being required to provide even moderately more specific statements would make employers more careful when contributing them and give consumers a greater ability to refute them.

The following comments include two expert reports that describe the problems with a consumer report form frequently used in the trucking industry. These experts suggest methods for easily modifying this form to greatly improve its accuracy. OOIDA also endorses two prior efforts in the federal government for improving the accuracy and usefulness of information. The first is the Equal Employment Opportunity Commission's Uniform Guidelines for Employee Selection Procedures.[2] The next is the Office of Management and Budget's "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies."[3] OOIDA's focus in these comments is the trucking industry. But the principles of accuracy embraced by these resources are applicable to all consumer reports.

III. COMMENTS

When the accuracy of a credit report regarding financial transcripts is called into question, it is rare for its precision to be at issue. The use of dollars and dates to describe a debt provides a universally recognized level of precision. "The Titanic cost $7.5 million to build and then sank on April 15, 1912." Because these statements contain precise expressions of a cost

[2] 43 FR 38295 (August 25, 1978)

[3] 67 FR 8452 (February 22, 2002)

and a date, an assessment of their accuracy would focus on their truth.

In consumer reports containing a person's employment history, however, such precise terms are not always used. For example, "Captain White's performance was less than satisfactory and he is ineligible for rehire by White Star Lines." The reader may not know what makes the Captain satisfactory or not. (Not everyone may know who skippered the Titanic). The reader does not know why the Captain is not eligible for rehire - because of his unsatisfactory performance, or because he went down with the ship?

In consumer reports containing statements related to the quality of a person's work, the units of measurement may not be as clear cut as financial or temporal data. How accurate is a statement that a person's work was "satisfactory?" By what standard or scale? If a person was said to "violate company policy," how would the user of such a consumer report know the seriousness of the violation or culpability of the employee labeled as such? Precision is not built into those phrases. Therefore, standards for the accuracy of statements provided to consumer reporting agencies must pertain to more than the transfer of data. It must also embrace both the truth and precision of its expression.

The precision problem is primarily one of definition. In the expert report attached as Exhibit 2, Dr. Ed Schiappa,[4] a well regarded expert in definitions, examines the need for the terms used in consumer reports to have sufficient definition to convey meaning. Dr. Schiappa states that meaning is critical for a word or phrase to be used consistently by different people. According to the U.S. Court of Appeals for the Tenth Circuit in its review of one credit reporting system, without sufficient meaning, a statement on an employment history form will not be used

[4] The identity of the subject consumer reporting agency has been redacted from the attached expert reports.

consistently by different employers and its accuracy will be in doubt.[5]

The task of assuring the accuracy of employment histories is not a difficult one. Social scientists have long embraced the concepts of validity and reliability for assessing the accuracy of a collection of information. The report of Dr. Glenn Beamer, attached as Exhibit 3, describes how collections of information, such as those routinely collected by consumer reports, can be evaluated for their reliability. The measurement of reliability is a concept that is embraced by scientists across many disciplines to evaluate the consistency of the data gathered. Furthermore, Dr. Beamer describes the need for employers to use valid forms to submit accurate employment information. Dr. Beamer's validity analysis refers to the structure and content of the questions asked on such forms. Dr. Beamer suggests several ways to improve the validity of the forms used to collect employment histories of truck drivers.

Drs. Schiappa and Beamer's analyses demonstrate how precision can be incorporated into statements about employment histories. As an example of an invalid collection of information, a box labeled satisfactory that may or may not be selected by an employer to describe an employee has little precision. If the box is not checked, the reader does not know whether the employer forgot to check the box or whether the driver was unsatisfactory. What level of performance makes a driver satisfactory to that employer? A five point scale from very unsatisfactory to very satisfactory, with each possible response defined, is more precise. The reader has more information on which to make a judgment and can better compare employee candidates measured with the same scale.

In support of these standards of reliability and validity, OOIDA submits, attached as

[5] *Cassara v. DAC Services*, 276 F.3d 1210, 1219-20 (10th Cir. 2002)

Exhibit 4, a copy of the Office of Management and Budget's "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies." In these guidelines, OMB recognized the responsibility of the government, in its position of authority and public trust, to take certain care in the production of information produced to the public. Such responsibility is directly analogous to the grave responsibility required by the Fair Credit Reporting Act of those who trade in the personal information and reputation of American consumers.

Finally, attached as Exhibit 5, OOIDA submits a copy of the Equal Employment Opportunity Commissions' Uniform Guidelines for Employee Selection Procedures. These guidelines were created to help employers establish objective valid employee screening devices. OOIDA understands that the Fair Credit Reporting Act does not dictate what information must be contained in a consumer report. But whatever statements are included must be amenable to an accuracy analysis. The EEOC Guidelines provide employment-specific examples of how consumer reports containing employment histories can be written in a way that improve their precision, objectivity, utility, and, therefore, accuracy.

IV. CONCLUSION

OOIDA believes that a consumer reporting system that produces inaccurate employment histories fails to serve both drivers and employers. Under such a system, good drivers can be forced to leave the industry, and bad drivers are allowed to remain. The form and content of statements provided by employers to consumer reporting agencies is significantly determined by the type of input device provided by the consumer reporting agency. This rulemaking should work to ensure that businesses only provide information or statements that are both truthful and

sufficiently precise to achieve maximum possible accuracy, no matter what the subject area.

Thank you for your consideration of these comments.

Respectfully submitted,

s\Paul D. Cullen, Jr.

JAMES J. JOHNSTON
President
Owner-Operator Independent
Drivers Association, Inc.

PAUL D. CULLEN, JR.
The Cullen Law Firm, PLLC
1101 – 30th Street, N.W. Suite, #300
Washington, DC 20007
(202) 944-8600

May 22, 2006

[redacted]

TERMINATION RECORD FOR CDL DRIVERS

1 MEMBER I.D. #

2 INDIVIDUAL'S LAST NAME | FIRST NAME INITIAL

3 SOCIAL SECURITY # | 4 DATE OF BIRTH Mo. Day Yr

5 PERIOD OF SERVICE TO

6 Was the driver involved in DOT or Non-DOT recordable accidents/incidents during this period of service? Yes ☐ No ☐ *If yes, please complete accident and/or drug alcohol information on reverse.*

7 Do you have record of the driver violating DOT drug/alcohol regulations in the past 3 years? ☐ ☐

8 STATE LICENSE #

STATE LICENSE #

Important Notice: Refer to "Guide" for full explanation of codes below (Form SL0050g)

9 **Eligible for Rehire** *(Circle only one)* 001 Yes 002 Yes, but against company policy 003 No 004 Review required before hiring

10 REASON FOR LEAVING
(Circle Only One)
- 101 Discharged (or Company Terminated Lease)
- 104 Agency Lease Terminated
- 106 Laid Off (or Lease Suspended)
- 112 Leave of Absence
- 122 Repossession/Lease Default
- 127 Retired
- 133 Resigned/Quit (or Driver Terminated Lease
- 199 Other _____

11 STATUS
(Circle All That Apply)
- 202 Company Driver
- 207 Lease Driver (Employee of Independent Contractor)
- 213 Owner/Operator
- 217 Lease Purchase Program
- 228 Trip Leaser
- 230 Student/Trainee
- 233 Student CDL Permit
- 234 Casual Driver
- 299 Other _____

12 DRIVER'S EXPERIENCE
(Circle All That Apply)
- 303 Local
- 305 Regional
- 311 Mountain Driving
- 327 Over the Road
- 332 Single Driver
- 333 Driver Trainer/Instructor
- 351 1st Driver on Team
- 352 2nd Driver of a Team
- 355 Freight Handling
- 399 Other _____

13 EQUIPMENT OPERATED
(Circle All That Apply)
- 505 Auto Transporter
- 511 Bus
- 516 Double Trailer
- 523 Driveaway/Towaway
- 527 Dry Box
- 529 Dump Truck
- 532 Flat Bed
- 533 Mobile Crane
- 534 Pick Up or Hot Shot
- 540 Refrigerated
- 542 Specialized Trailer
- 544 Specialized Truck/Toter
- 547 Straight Truck
- 549 Pneumatic Trailer
- 552 Tank Truck
- 562 Triple Trailer
- 573 Van
- 581 Winch
- 599 Other _____

14 LOADS HAULED
(Circle All That Apply)
- 707 Bulk Commodity
- 712 Container
- 713 Empty Trailer
- 714 Gen. Commodity
- 716 Electronics
- 718 Hanging Meat
- 720 Hazardous Material
- 725 Household Goods
- 729 Livestock
- 730 Lumber
- 731 Machinery
- 733 Mobile Homes
- 735 Motor Vehicle
- 750 Passengers
- 762 Oversized Loads
- 763 Parcels
- 764 Pipe
- 769 Refrigerated
- 773 Steel
- 799 Other _____

15 WORK RECORD
(Circle All That Apply)
- 901 Satisfactory
- 902 Superior
- 903 Outstanding
- 912 Excessive Complaints
- 913 Cargo Loss
- 915 Falsified Employ. Application
- 917 Equipment Loss
- 924 Late Pick Up/Delivery
- 926 Log Violation
- 928 No show
- 929 Failed To Report Accident
- 931 Quit Under Dispatch/Did Not Possess a Load
- 933 Quit/Dismissed During Training/ Orientation/Probation
- 935 Company Policy Violation
- 938 Unsatisfactory Safety Record
- 940 Disconnected Tracking Device
- 944 Personal Contact Requested
- 957 Unauthorized Equip. Use
- 959 Unauthorized Passenger
- 961 Unauthorized Use of Company Funds
- 999 Other _____

QUIT UNDER LOAD/ABANDONMENT
(If Applicable, Circle Only One)
- 950 Co. Terminal – With Notice
- 951 Auth. Location – With Notice
- 952 Co. Terminal – W/O Notice
- 953 Unauth. Location – W/O Notice
- 954 Left Vehicle With Team Driver
- 955 Unauth. Location – With Notice
- 956 Auth. Location – W/O Notice

16 _____
List Disputed Employment Code(s) separated by a comma

17 _____ _____
Contact Person (optional) *Date*

[redacted]

TERMINATION RECORD DETAILS

[A] MEMBER I.D. # [B] SOCIAL SECURITY #

Please detail all DOT recordable accidents (use additional paper if more than 8) and the most severe non-DOT recordable accidents

[C] _____ Total Number of DOT Recordable Accidents [D] _____ Total Number of Non-DOT Accidents/Incidents

[E] Accident/ Incident#	[F] DOT?	[G] Date	[H] City	[I] State	[J] #of Injuries	[K] #of Fatalities	[L] HAZMAT	Description Codes(s) (Use up to 4)
Sample	Y or N	3/3/1998	Tulsa	OK	0	0	Y or N	01, 02
1	Y or N						Y or N	
2	Y or N						Y or N	
3	Y or N						Y or N	
4	Y or N						Y or N	
5	Y or N						Y or N	
6	Y or N						Y or N	
7	Y or N						Y or N	
8	Y or N						Y or N	

ACCIDENT DESCRIPTION CODES

01 Backing	09 Overturn	17 Picked Up Damaged Trailer	24 Fuel Spill	40 Preventable Accident/ Incident
02 Right Turn	10 Struck Overhead Object	18 Mechanical Failure	25 Fire	50 Non-Preventable Accident/Incident
03 Left Turn	11 Jack Knife	19 Ran Traffic Control	26 Hit Animal	99 Misc.
04 Lane Change Side Swipe	12 Hit Pedestrian	20 Ran Off Roadway	27 Hit Parked Vehicle	
05 Rear End Collision	13 Passing	21 Downgrade Runaway	28 Load shift	
06 Intersection Collision	14 Dropped Trailer	22 Trailer Breakaway	29 Theft	
07 Head On Collision	15 Hit While Parked	23 Cargo Spill	30 Rollaway	
08 Struck Stationary Object	16 Hit While Moving		31 Non Contact	

[M] List disputed accident/incident number(s): (1,2,3 etc.) and/or items C and D _____

Indicate below any violations of 49 C.F.R. Part 40 that occurred during this period of service.

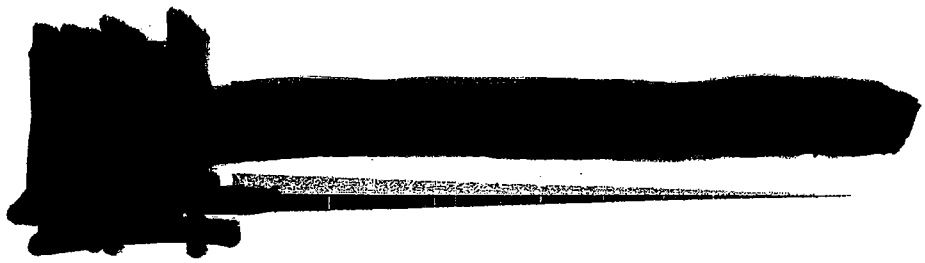
Violation		Date(s) of violation: Mo Day Yr	Mo Day Yr
1	☐ Driver had an alcohol test with a confirmed B.A.C. of 0.04 or greater.		
2	☐ Driver had a controlled substance test with a positive result.		
3	☐ Driver refused a controlled substance or alcohol test (includes verified adulterated or substituted results).		
4	☐ Driver violated other DOT drug/ alcohol regulations.		

Indicate below any violations of 49 C.F.R. Part 40 that occurred within the past 3 years which were reported to you by a previous employer.

Violation		Date(s) of violation: Mo Day Yr	Mo Day Yr
5	☐ Previous employer reported DOT drug/alcohol violation(s).		

CHECK HERE ☐ if driver disputes Drug/Alcohol violation(s).

[redacted]



GUIDE TO THE TERMINATION RECORD FORM

This guide includes the definitions of codes and terms used in the current version of [redacted] Termination Record Form for CDL drivers.

While most of the descriptions on the termination form are self-evident, we recommend you consult this guide when contributing termination records if you are in doubt.

You may also receive employment histories contributed by other [redacted] members that contain terms not used in your operation. Use this guide to interpret any term in which you are unsure of the meaning.

TERMINATION RECORD DETAILS

1 MEMBER ID #: Record the customer number assigned by [redacted]

2 LAST NAME AND FIRST INITIAL OF FIRST NAME: Record driver's last name with no space and no punctuation. In the box at the end of this line, record the first initial of the driver's first name.

3 SOCIAL SECURITY NUMBER: Record the driver's social security number.

4 DATE OF BIRTH: Record the month, day and year (including century) of the driver's date of birth.

5 PERIOD OF SERVICE: Record period of service using starting month and year (including century) to terminating month and year (MMCCYY to MMCCYY).

6 WAS THE DRIVER INVOLVED IN DOT OR NON-DOT ACCIDENTS/INCIDENTS DURING THIS PERIOD OF SERVICE?: Answer by marking the yes or no box. If yes, give accident details on the back of the form.

7 DO YOU HAVE RECORD OF THE DRIVER VIOLATING DOT DRUG/ALCOHOL REGULATIONS IN THE PAST 2 YEARS?: Answer by marking either the yes or no box. If yes, indicate the violation on the back of the form.

8 STATES OF LICENSE: Record the post office abbreviation for the state or states in which the driver has held licenses while with your company. Record license number(s) omitting all spaces and dashes.

9 ELIGIBLE FOR REHIRE (circle only one code)
001 Yes: Driver is eligible for rehire.
002 Yes, but against company policy: Driver is qualified, but your company has a policy against rehiring drivers regardless of qualifications.
003 No: Driver is ineligible for rehire based on current company standards.
004 Review required before rehiring.

10 REASON FOR LEAVING (circle only one code)
101 Discharged: Employment or lease is involuntarily terminated.
104 Agency Lease Terminated: An agency affiliated with the company has terminated, closed, or is no longer under contract.
106 Laid Off: Driver is laid off or lease has been suspended due to business reasons unrelated to performance.
112 Leave of Absence: Company approved leave without pay.
122 Repossession/Lease Default: Owner Operator/Independent Contractor has defaulted on a lease contract or had their truck(s) and/or trailer(s) repossessed.
127 Retired: Driver retires.
133 Resigned/Quit: Employment or lease is voluntarily terminated.
199 Other: Anything other than items listed above. This space is provided for your documentation. [redacted] will record "other" only.

11 STATUS (You may circle more than one code)
202 Company Driver: An employee of the company.
207 Lease driver: Employee of an independent contractor.
213 Owner/operator: A person who owns and drives his own equipment for a company as its employee or as an independent contractor.
217 Lease Purchase Program: A driver that is currently participating or has participated in an equipment lease purchase program.
233 Student CDL Permit: A student qualified as a second seat driver while on a CDL learner permit.
228 Trip Leaser: Driver is acting as an independent operator or as an agent of a carrier contracting with your company for specific loads hauled on a trip by trip basis.
230 Student/Trainee: A student or trainee of the company.
234 Casual Driver: A driver hired to drive on an intermittent, casual, or occasional basis who may or may not be an employee of the company.
299 Other: Anything other than items listed above (see 199).

12 DRIVER'S EXPERIENCE (You may circle more than one)
303 Local: Driver had substantial city driving experience.
305 Regional: Driver had substantial regional driving experience.
311 Mountain driving: Driver had substantial mountain driving experience.
327 Over-the-Road: Driver had substantial long haul driving experience.
332 Single Driver: Driver had sole responsibility for equipment and substantial experience driving alone.
333 Driver Trainer/Instructor: Driver had substantial road experience training students and/or trainees. A company employee that has a substantial amount of experience with classroom and driving instruction.
351 1st driver of a Team: Driver had primary responsibility in a two-member team.
352 2nd driver of a Team: Driver had secondary responsibility in a two-member team.
355 Freight Handling: Driver had substantial experience loading and unloading freight.
399 Other: Anything other than items listed above (see 199).

13 EQUIPMENT OPERATED (You may circle more than one)
505 Auto Transporter: Truck, semi-trailer, or trailer with the body designed for the transportation of other vehicles.
511 Bus: A motor vehicle designed, constructed and used for the transportation of passengers.
516 Double trailer: (Also twin trailer-unit) consists of tractor, semi-trailer and full trailer.
523 Driveaway/Towaway: Motor vehicle(s) or trailer(s) constitute the commodity being transported. One or more sets of wheels of such vehicles are on the road during transportation.
527 Dry box: Enclosed semi-trailer.
529 Dump truck: Truck, semi-trailer or trailer which can be tilted to discharge load.
532 Flat bed: Truck or trailer without sides or top.
533 Mobile Crane: A truck designed for the specific purpose of transporting a crane.
534 Pick-Up or Hot Shot:: Up to one ton truck with or without a trailer.
540 Refrigerated: Refrigerated truck or trailer designed for hauling perishables.
542 Specialized trailer: A trailer designed for a specific purpose not included in the other categories listed (e.g. missile carrier).
544 Specialized truck/Toter: A straight truck/tractor with the body designed for a specific purpose other than those listed in other categories here (e.g. concrete, refuse, etc.).
547 Straight truck: A truck with the body and engine mounted on the same chassis and not listed elsewhere under equipment operated.
549 Pneumatic Trailer: Truck, semi-trailer or trailer loaded and/or unloaded using compressed air.
552 Tank truck: Truck, semi-trailer, or trailer with a tank body for hauling petroleum, chemicals, liquids, or dry commodities in bulk.

[redacted]

562 Triple trailer: Tractor, semi-trailer plus two trailers.
573 Van: Van, including step van.
581 Winch: Hoist used on straight truck or tractor (includes gin pole).
599 Other: Anything other than items listed above (see 199).

14 LOADS HAULED (You may circle more than one)
707 Bulk Commodity: Liquid or dry bulk.
712 Containers: Hauling of large cargo-carrying containers that can be easily interchanged between trucks, trains, and ships, without rehandling contents.
713 Empty trailer: Driver delivers empty trailers–does not apply to deadheading.
714 General Commodity: Varied types of freight.
716 Electronics: Transporting electronic commodities requiring special handling.
718 Hanging meat: Self explanatory.
720 Hazardous material: As designated by the Department of Transportation including but not limited to: explosives, radioactive materials, etiologic agents, flammable liquid or solids, combustible liquids or solids, poisons, oxidizing or corrosive materials, and compressed gases.
725 Household goods: Self explanatory.
729 Livestock: Transporting cattle, horses, etc.
730 Lumber: Self explanatory.
731 Machinery: Self explanatory.
733 Mobile homes: Self explanatory.
735 Motor vehicles: Transporting of motor vehicles by hauling them on special vehicles or through driveaway-towaway.
750 Passengers: People.
762 Oversized loads: Loads requiring special permits due to size or weight.
763 Parcels: Parcels and packages.
764 Pipe: Self explanatory.
769 Refrigerated: Self explanatory (not including hanging meat).
773 Steel: Other than pipe.
799 Other: Anything other than items listed above (see 199).

15 WORK RECORD (You may circle more than one)
It is strongly recommended that items denoting less than satisfactory performance be supported by documentation in the driver's file.
901 Satisfactory: Driver meets minimum company standards of performance in all categories.
902 Superior: Driver exceeds minimum company standards of performance in all categories.
903 Outstanding: Driver's performance is outstanding in all categories.
912 Excessive Complaints: An excessive number of complaints have been received regarding the driver's service and/or safety.
913 Cargo loss: Cargo was lost, stolen, damaged or destroyed while assigned or under direct responsibility of the driver.
917 Equipment loss: Equipment was lost, stolen, damaged or destroyed while assigned to or under direct responsibility of driver.
915 Falsified Employment Application: Falsified information on employment application or omitted information as required by company, state, or federal regulations.
924 Late pick up/Delivery: Failed to make pickup or delivery according to schedule.
926 Log Violation: Violation of Federal Motor Carrier Safety Regulations, "Hours of Service," part 395.
928 No show: Driver failed to appear on job site without notification or approval of supervisor. Driver has hauled previous loads for the company.
929 Failed To Report Accident: Driver violated accident reporting requirements while in the service of the company.
931 Quit Under Dispatch: Driver was available for work, assigned a load but quit before load was secured. Driver did not possess a load.
933 Quit/Dismissed During Training/Orientation/Probation: Driver did not complete company training, orientation and/or probabtion. If the driver quit or was dismissed during orientation, leave sections 12, 13 & 14 blank and do not provide further information to section 15.
935 Company policy violation: Driver violated company policies and/or procedures. Use this code only if the other selections in this section do not indicate the company policy violated.
938 Unsatisfactory Safety Records: Driver did not meet company safety standards.
940 Disconnected Tracking Device: The driver disconnected the truck and/or trailer-tracking device(s) without company authorization
944 Personal Contact Requested: Company issuing record has further information to provide regarding the driver or for the driver.
957 Unauthorized equipment use: Deviated from route or used equipment for purposes not specified by company. (Not intended to be used when the driver has resigned/quit or terminated lease and returned equipment to the nearest company terminal or a location authorized by the company.)
959 Unauthorized passenger: Passenger in company vehicle contrary to company policy or did not meet company policy requirements covering authorized passenger.
961 Unauthorized Use of Company Funds: Driver used company funds for purposes not authorized by company.
999 Other: Anything other than items listed above (see 199).

QUIT UNDER LOAD/ABANDONMENT: (Circle only one code, if applicable) Quit job before truck and/or cargo was delivered to final destination.
950 Co. Terminal - With Notice: Left truck and/or cargo at a company terminal. Driver did notify the company of termination. (Not intended to be used when the driver has resigned/quit or terminated lease and returned equipment to the nearest company terminal or a location authorized by the company.)
951 Auth. Location - With Notice: Left truck and/or cargo at a location authorized by the company. Driver did notify the company of termination.
952 Co. Terminal - W/O Notice: Left truck and/or cargo at a company terminal. Driver did not notify the company of termination. (Not intended to be used when the driver has resigned/quit or terminated lease and returned equipment to the nearest company terminal.)
953 Unauth. Location - W/O Notice: Left truck and/or cargo at a location unauthorized by the company. Driver did not notify the company of termination.
954 Left Vehicle With Team Driver: Left truck and/or cargo in the possession of a team driver.
955 Unauth. Location - With Notice: Left truck and/or cargo at a location unauthorized by the company. Driver did notify the company of termination.
956 Auth. Location - W/O Notice: Left truck and/or cargo at a location authorized by the company. Driver did not notify the company of termination.
**** The following codes are no longer used, but could appear on older termination records.
909 Abandonment: Abandoned truck and/or cargo without notification to the company.
937 Quit Under Load: Quit job before truck and/or cargo was delivered to final destination. Assumes that driver did notify company of termination.

16 LIST DISPUTED EMPLOYMENT CODES(S): List any employment codes that were disputed by the driver at the time of termination.

17 CONTACT PERSON: Signature of individual completing form and date.

[redacted]

A Member I.D.#: Record the customer number assigned by [redacted]

B Social Security #: Record the driver's Social Security Number.

ACCIDENT/INCIDENT DETAILS

C Total Number of DOT Recordable Accidents: Record the total number of accidents that are classified as "recordable" under DOT guidelines. The number of accidents listed does not necessarily reflect fault on the part of the driver involved.

D Total Number of Non-DOT Accidents/Incidents: Record the total number of accidents/incidents that do not meet the DOT recordable classification. Accidents/incidents listed do not necessarily reflect fault on the part of the driver involved.

E DOT?: Circle Y or N to indicate whether the accident meets the DOT guidelines for a recordable accident or did not meet those guidelines.

F Date: Record the month, day and year (including century) the accident/incident occurred.

G City: Record the city or town (or nearest) in which the accident/incident occurred.

H State: Record the state in which the accident/incident occurred.

I Injuries: Record the total number of persons injured as a result of the accident/incident who immediately received medical treatment away from the scene.

J Fatalities: Record the total number persons that died as a result of the accident/incident.

K HAZMAT: Record whether hazardous materials, other than fuel from the tanks of motor vehicles involved in the accident/incident, were released.

L Description Code(s): You may use up to four (4) accident description codes to describe each accident.

01 Backing: Occurred while backing.
02 Right Turn: Occurred while making a right turn.
03 Left Turn: Occurred while making a left turn.
04 Lane Change Side Swipe: Involved a side swipe collision while changing lanes.
05 Rear End Collision: Involved striking another vehicle in the rear.
06 Intersection Collision: Involved a collision while in an intersection.
07 Head-On Collision: Involved a head-on collision with another vehicle.
08 Struck Stationary Object: Involved an impact with a stationary (fixed) object.
09 Overturn: Involved the truck and/or trailer overturning.
10 Struck Overhead Object: Involved an impact with an overhead object.
11 Jack-knife: Involved the truck and trailer jack-knifing.
12 Hit Pedestrian: Involved an impact with a pedestrian.
13 Passing: Occurred while passing another vehicle.
14 Dropped Trailer: Involved the trailer being dropped.
15 Hit While Parked: Involved the truck and/or trailer being hit while parked.
16 Hit While Moving: Involved the truck and/or trailer being hit by another vehicle or object while moving.
17 Picked Up Damaged Trailer: The driver picked up a trailer that had been previously damaged.
18 Mechanical Failure: Occurred due to mechanical failure of the truck and/or trailer.
19 Ran Traffic Control: Involved a failure to yield at traffic control.
20 Ran Off Roadway: Involved the truck and/or trailer running off the roadway.
21 Downgrade Runaway: Involved loss of control on a downgrade.
22 Trailer Breakaway: Involved the trailer breaking away from the tractor.
23 Cargo Spill: Involved a cargo spill.
24 Fuel Spill: Involved a fuel spill from the power unit.
25 Fire: Involved a fire.
26 Hit Animal: Involved an impact with an animal.
27 Hit Parked Vehicle: Involved an impact with a parked vehicle.
28 Load Shift: Involved a load shift.
29 Theft: Involved theft of the truck, trailer and/or cargo.
30 Rollaway: Involved a parked truck and/or trailer rollaway.
31 Non-Contact: Did not involve a collision.
40 Preventable Accident/Incident: Based on your company guidelines the accident/incident was preventable.
50 Non-Preventable Accident/Incident: Based on your company guidelines the accident/incident was non-preventable.
99 Misc.: Anything other than the items listed above.

M List Disputed Accident/Incident Number(s): List the accident/incident number(s) (1, 2, 3, etc. in the preceding chart) that the driver disputes. If the driver disagrees with the total number of DOT or non-DOT accidents/incidents, enter items C and/or D.

DRUG/ALCOHOL VIOLATION DETAILS

1 Driver had an alcohol test with a confirmed B.A.C. of 0.04 or greater: If the driver violated this section of 49 C.F.R. Part 40, Mark an "X" in the box and list the date(s) of violation.

2 Driver had a controlled substance test with a positive result: If the driver violated this section of 49 C.F.R. Part 40, mark an "X" in the box and list the date(s) of violation.

3 Driver refused a controlled substance or alcohol test (includes verified adulterated or substituted results): If the driver violated this section of 49 C.F.R. Part 40, mark an "X" in the box and list the date(s) of violation.

4 Driver violated other DOT drug/alcohol regulations. If the driver violated this section of 49 C.F.R. Part 40, mark an "X" in the box and list the date(s) of violation.

5 Previous employer reported DOT drug/alcohol violation(s). If the driver violated this section of 49 C.F.R. Part 40, mark an "X" in the box and list the date(s) of violation.

Check Here: Mark an "X" in the box if the driver disputes any of the drug/alcohol violations. [redacted]

Report on Definitional Issues in [redacted]

Edward Schiappa, University of Minnesota

June 2, 2005

Executive Summary

- The purpose of definitions is to provide precise, accurate meanings to a word or phrase. Good definitional practices facilitate *denotative conformity* (agreement about what a word or phrase refers to) and *connotative predictability* (a reliable sense of the reactions a word or phrase elicits).

- Good definitions or category descriptions provide clear exemplars of the phenomenon being defined such that members of a particular language community understand that "X counts as Y in context C." Members of a language community, such as the trucking industry, must have shared understanding of how information is *encoded* into data and how data should be *decoded* accurately.

- Good definitional practices meet four criteria: Clarity, Shared Purpose, Appropriate Authority, and Feedback. Collectively, these practices facilitate a language community's shared understanding of what "attributes" are central and important to the catetgories used by that community.

- The following phrases and definitions used in [redacted] Termination Record Form and [redacted] *Guide to the Termination Record Form* were analyzed: "company policy violation," "unsatisfactory safety record," "excessive complaints," "cargo loss," "equipment loss," "quit/dismissed during training/orientation/probation," "eligible for rehire: no," "other," "personal contact requested," "late pick up/delivery," "log violation," "no show," "failed to report accident," "quit under dispatch," "unauthorized equipment use," "unauthorized passenger," and "unauthorized use of company funds."

  In all cases, the "definitions" provided were seriously flawed: They were circular, vague, ambiguous, or open to abuse. They fail to facilitate denotative conformity or connotative predictability.

- The definitional practices of [redacted] fail to meet the four criteria of good definitional practices. The flawed design of the code categories can be understood clearly by considering how the codes *could* have been defined more clearly.

- Accurate interpretation of data generated by TRF reports is impossible. The problems are systemic to the design of the form and its definitional glossary. The Work Record section of the TRF does not meet the goal of "maximum possible accuracy of the information concerning the individual about whom the report relates."

Report on Definitional Issues in [redacted]

Edward Schiappa, University of Minnesota

June 2, 2005

This report is divided into three sections. Section I describes a set of standards for understanding and evaluating definitions and categories. Section II provides an analysis of the definitions provided in the *Guide to Termination Record Forms* distributed by [redacted] Section III provides an overall assessment of the definitional issues.

I. Standards for Definitions & Categories

In this section I provide a set of criteria for evaluating definitions and categories. I frame my remarks as answers to a series of questions: What is a definition? What is the difference between a definition and a category (or "classification")? What is the purpose of definition? And, lastly, What are the criteria for good definitional practices?

What is a definition?

Since definition is a topic that has been of interest for well over 2,000 years, it is not surprising that there are actually a number of definitions of "definition" (Robinson, 1950; Rey, 2000). Aristotle is credited for the standard definitional form involving genus and difference: An *X* is (a kind of) *class name* that has such-and-such *attributes*. I will discuss categories and attributes in the following subsection. Before that discussion, we need to recognize that distinctions are drawn among lexical, ostensive, operational, theoretical, stipulative, circular, and other types of definition. It is not necessary to discuss all of these types of definition, but four are particularly relevant. First, a *lexical* definition is simply the sort of definition found in a

dictionary. It is an empirical guide to usage; that is, a dictionary tells us what the most common use of words has been, and thus functions as a prescriptive guide for how language users should use the word now.

For ordinary, day-to-day use, a standard dictionary is adequate. Groups of language users often have specific needs and interests that require them to use words in a more precise way than is common in ordinary language use. Obvious examples of this would be legal, medical, and scientific terms. In such specialized language communities, a good deal of effort is expended defining words in a precise manner. Ordinary words take on a far more specific meaning within a specialized language community (such as "force" in physics or the law). It should be noted, however, that it is not only the highly specialized fields of law, medicine, and science that develop their own special uses for words. Indeed, any time an identifiable group of people share a common set of experiences, they can be described as a language community that develops a particular set of language practices that mark them as distinct. If a person becomes a musician, an auto mechanic, a professional poker player, a salesperson, or a truck driver, part of learning how to be part of that community involves learning to "talk the talk." Joining a community, such as "the trucking industry," is joining a *language* community that uses words in a particular fashion. Some of those words may be unique to that community, and other words may be taken from ordinary usage but given more specific meaning within that community.

For specialized language communities, reliance on lexical definitions is not enough. There is a need for what are called "stipulative" and "operational" definitions. A *stipulative* definition is simply a declaration and agreement by a language community that a word "Y" will be used in a particular fashion. Whomever first called the manual graphical user interface part of a computer a "mouse" simply declared it to be so, and now everyone knows what we are talking

about when we refer to a computer's "mouse"—even though that use is obviously quite different than the traditional lexical definition of a mouse. Furthermore, when it is important to have common agreement about when something should be called "Y," we often develop an *operational* definition. An operational definition often specifies some *measurable* dimension. In education, "gifted" and "challenged" are often defined by reference to a specific score on a standardized intelligence test. Many psychological diagnoses are dependent on specific scores measured by detailed questionnaires. Vehicles are often categorized by such measurable dimensions as weight, size, and number of tires. "Speeding" is operationalized by travel at a speed in measurable excess of posted limits.

Lastly, it is important to note that what is called a "circular definition" is *not* an acceptable form of definition. A circular definition is one that simply repeats the word or phrase being defined in the definition itself without providing additional information about the word or phrase's denotative or connotative meaning. Since circular definitions assume a prior understanding of the word or phrase being defined, it does not provide members of a language community any insight into how the word or phrase should be used.

To summarize: Specific language communities develop, through practice over time or through concrete acts of stipulation, general and operational definitions that guide the linguistic behavior of the community's members. What these definitions have in common is a desire for clear and consistent use of specific words. Formally, they create a linguistic "rule" of the form "X counts as Y in context C." Thus, a "flush" in poker *counts* as a flush only if one has a sufficient number of cards of the same suit. That use of the word "flush" is obviously quite different than how the term might be used by a plumber or doctor. Accordingly, the same word

might be defined quite differently by different language communities, depending on their respective needs and interests.

For the purposes of this report, "definition" refers to a specific effort by a language community to identify the denotative and connotative meanings of a word. What I wish to stress at this point is that a definition functions within a language community as a kind of linguistic rule, "X counts as Y in context C."

What is the difference between a definition and a category?

Much of what I have said so far about definitions could also be said about "categories" and systems of "classification." As communication scholars Bowker and Star note, "to classify is human" (1999). Stressing the importance of categorization, Senft (2000) argues, "classification abilities are necessary to the survival of every organism" (p. 11). Similarly, Bowerman notes "the grouping of discriminably different stimuli into categories on the basis of shared features is an adaptive way of dealing with what would be an overwhelming array of unique experiences" (1976, pp. 105-6). In short, the way we make sense of the world is through the aquisition of categories. This is also a useful way to think about how language works—primarily as a complex system of categories used to make sense of an infinitely complex world.

Categories are formed based on learning the relevant functional, perceptual, or other sorts of attributes that members of a category share. This is precisely why Aristotle's formulation of definitions is so influential: An _X_ is (a kind of) *class name* that has such-and-such *attributes*. "Attributes" are simply features or qualities of a phenomenon: a chair is something we sit on (a functional attribute), a ball is round (a perceptual attribute, something we see). One's earliest exposure to a category is sometimes called an original or prototypical exemplar (Bowerman,

1976). It is through exposure to a series of examples (or "exemplars") that we learn what counts as a member of a category. One typically does not learn what a "ball" is from one example, since balls have attributes that other categories have as well (not all round objects are balls). For a category to be meaningful and useful, it must both include items and exclude others, thus humans acquire a social category by learning a set of "similarity/difference relationships" that distinguish one category from another (Schiappa, 2003). We have to learn when something "counts" as a member of *this* category but not *that* one, and we do that by learning what attributes one category has in common that are *different* from the attributes of another category. Some linguists and philosophers refer to this process as "semantic mapping." That is, we must learn how our words map out the world around us, and we must learn to "read" that map in a manner consistent with other members of our language community: "A network of definitions maps experience by categorizing" (Matthews, 1998, p. 55).

The production of definitions is a social practice designed to *formalize* our understanding of specific categories. Definitions identify the "definitive" or "essential" attributes that characterize a category. Definitions are ultimately intended to serve a social purpose of *stabilizing* meaning so that when a person refers to a category, we know what that person is talking about.

What is the purpose of definitions?

Though I have already said that definitions serve an important stabilizing function so that we can understand each other, especially in specialized language communities, a few additional remarks may be useful to understand the purpose that definitions have. The key idea is that definitions are intended to have more *precise* and *predictable meaning* than mere "description":

Descriptions "do not constrain experience as a network of definitions do. Descriptions are open-ended" (Matthews, 1998, p. 56). To explain how definitions function more precisely and predictably than descriptions, I next describe the concepts of "denotative conformity" and "connotative predictability."

Denotative conformity refers to the degree of intersubjective agreement about what a specific word *refers* to. To "denote" means to "refer," to point out something, as in "there's a tornado!" Denotative conformity can be measured. For example, among experienced poker players, one would find 100% agreement about what the terms "flush" and "straight" refer to. The degree of denotative conformity varies among different language communities. A term like "solenoid" might have relatively low denotative conformity among a general population (I would *not* know one if I saw it, for example), but it would undoubtedly have a near perfect degree of agreement among experienced mechanics.

Connotative predictability is similar, but refers to the subjective "sense" of a word rather than its objective referent. All words conjure up thoughts, including images, feelings, and attitudes. Sometimes those thoughts are mundane (such as the word "pencil"), and other times the feelings and attitudes elicited by a word can be quite powerful (such as the word "murder"). Part of what definitions help to do is to stabilize the connotative predictability of a word so that when person A uses a word, that person can predict the sorts of images, feelings, and attitudes person B will have in response. This is why politicians use highly charged words like "terrorist" or "freedom," of course, but the same principle would apply to almost any word used in a specific language community. If a veteran professional baseball player refers to another player as a "rookie," the term has both a denotative meaning (referring to a player in his first year of

major league play) and a set of predictable connotative meanings (inexperienced and eager, for example).

Definitions play a crucial role in the encoding/decoding process of communication. The concepts of encoding and decoding have been crucial parts of models of communication for over 50 years, most notably in the Osgood and Schramm model (Schramm, 1954) that stressed all communicators are "interpreters" who must encode and decode information. *Encoding* is the process of converting a complex set of information into more manageable "bits" of data. This is what language does: Words reduce an infinitely complex set of experiences into manageable and shareable chunks of information. However, such data or information are *meaningful* only if they are *decoded* accurately. *Decoding* is the reverse process of converting data that has been sent by a source into meaning (denotative and connotative) understandable by a receiver. Much of what we mean by learning to "talk the talk" of a particular language community involves learning to encode and decode in a manner consistent with veteran members of that language community, and here definitions can play an important role.

The bottom line purpose of definitions is *shared meaning*. Put simply, we want to know what a person *means* when he or she uses a word. Though "meaning" is a vexed term itself, all linguists and communication scholars certainly recognize the fundamental attributes of meaning include what Gottlob Frege described in 1892 as "sense" (connotative meaning) and "reference" (denotative meaning); that is, the subjective thoughts a word elicits in the mind of a hearer, and the objective referent to which a word refers.

Put more formally: The social goal of definition is to foster a coordinated and common understanding of words so that members of a language community have a high degree of denotative conformity when they use words to refer to the people, objects, and events most

relevant to that community, as well as connotative predictability so that they can anticipate the likely response to their use of such words. Similarly, "accuracy" in communication can be operationalized in the same fashion: To understand the meaning of a word "accurately" means that one understands its denotative reference and connotative sense with precision.

What makes for a good definition?

The proof of a good definition is in its performance. That is, if a particular language community defines a word such that its members recognize that X counts as Y in context C, then one should find a high degree of denotative conformity and connotative predictability. If a language community achieves high levels of denotative conformity and connotative predictability, it has a successful practice of definition. If not, then it does not have a successful practice of definition.

I would suggest four criteria that can assist in identifying successful definitional practices: Clarity, Shared Purpose, Appropriate Authority, and Feedback.

Clarity: As mentioned previously, we learn a category by being taught clear exemplars. By "clear exemplars" I mean examples that highlight the similarity/difference relations that distinguish one category from another. So, while not all birds can fly, one can learn the meaning of the category "bird" best through examples of birds that fly. There is clear evidence, for example, that a small child will learn to categorize "birds" better by initially being shown robins rather than penguins (Roberts & Horowitz, 1986). By contrast, one would not be advised to try to teach someone the meaning of the category of "chair" by first showing them a beanbag chair.

Learning a category involves learning what attributes are "essential" or "definitive" of a class of objects, events, or people. Thus, it would be preferable to learn who counts as an

"attorney" by reference to the attribute of "passing the bar exam" rather than, say, "someone who likes to argue." The first attribute is more essential or definitive than the second, and it helps differentiate between attorneys and non-attorneys more clearly.

Accordingly, the first criterion of a good definitional practice is that it strives for clarity through clear examples that allow members of a language community to recognize what the key attributes of a category are.

Shared Purpose: What counts as "essential" or "key" attributes of a category depends on members of a language community having a shared purpose in defining a given word. When I use the word "essential" I am not referring to some sort of metaphysical essence. Rather, I am referring to those attributes that the history and values of a given community deem as crucially important, given the community's shared purposes. Definitions are driven by needs, interests, and values. That is, we do not define words just for fun, but rather because of specific needs and interests that are reached when we have agreement on how to use certain words. For example, there are many ways to define "wetlands" and sometimes those definitions compete as government agencies and legislators have to decide what "counts" as a wetland within the meaning of specific laws and regulations. Ultimately, what is at stake is deciding what attributes (such as the presence of hydrophytes—plants that only grow in anaerobic conditions—versus how many days of the year there is standing water) are most important given the purposes of environmental protection laws.

It is unlikely that a language community will achieve *clarity* in its definitional practices unless it also has a common and *shared purpose* in defining important words. One cannot establish a clear category, with a clear set of definitive attributes, unless there is shared purpose. Without shared purposes for defining a word, it will be difficult if not impossible to agree on

what similarity/difference relations should be learned to know the rules for when X counts as Y in context C. In other words, a member of a language community cannot know if an X should count as a Y or not-Y without some understanding of the purpose of defining the category in the first place.

Appropriate Authority: An important criterion to consider when evaluating a set of definitional practices is who should have the *power* to define. When children are learning a language, it clearly advances the social interests of denotative conformity and connotative predictability to stipulate that parents and teachers have that power. When people are newcomers to a language community, such as medical students, law students, or apprentice laborers, it also makes sense that veterans have the authority and power to teach such newcomers what is what. In short, becoming a member of a language community involves initially "surrendering" definitional authority to those with more experience. As I said before, to be socialized into a particular community, one must learn to talk the talk.

Once one is socialized into a community, however, the question of how words should be defined is more a matter of negotiation and persuasion. For example, the faculty members of a new department might need to define what counts as a "scholarly publication" for the purposes of annually reviewing the achievements of each faculty member. Obviously, the department would want to achieve clarity in such a definition so that all faculty members would know what counts (denotative conformity) since scholarly publication is highly valued (connotative predictability). Through persuasion and negotiation, the department would identify what faculty members agreed were the most important attributes that should define the category, such as peer review and respected academic publishers. In such a case, the democratic norms of faculty

governance would be invoked since all faculty members would be recognized as authorized members of the language community.

Deciding who the appropriate authority should be in the practice of definition would vary from language community to language community. In the legal arena, the Supreme Court is the ultimate authority for defining what the words of the U.S. Constitution mean. In terms of deciding the definitions that appear in standard dictionaries, in a sense *everyone* is an appropriate authority because dictionaries are supposed to reflect what the most common uses of a word are.

I would suggest two ways to think about who the appropriate authority for defining should be. Ideally, *all* members of a specific language community share a stake in definitions. The best way to achieve denotative conformity and connotative predictability is to try to define terms as they are understood by all, or as many as possible, members of that community. Thus, just as in the case of dictionary definitions, the best way to foster the social goals of definition is through a "democratic" process that reflects the shared purposes of all members of that language community.

In cases where a "democratic" approach is not practical, such as a highly contested area of the law, definitional authority may have to be highly centralized. However, when such a circumstance obtains, the *other* criteria I have identified become all the more important. For example, if a group of faculty in a new department could not come to an agreement about how to define "scholarly publication," it could become necessary for a college dean to stipulate how scholarly publication will be defined for the purposes of reviewing faculty achievement. If that were to happen, it would be crucially important that the Dean meet the other criteria I have identified, including *clarity* and *shared purpose.* If the faculty members did not understand how the Dean defined scholarly publication, the group would risk not achieving their collective goals.

An individual faculty member might publish in an online, non-peer-reviewed journal, for example, then be outraged to learn after the fact that such an action does not "count" as scholarly publication.

In other words, regardless of who has the power to define, *all* members of a language community must be "empowered" with a clear understanding of the salient definitions of their community. Otherwise, the whole point of defining (denotative conformity and connotative predictability) is lost.

Feedback: An important part of how any word is learned is through the process of feedback. For example, small children will make mistakes of *overextension* (using a word too broadly, as in calling all round objects "balls") and *underextension* (not recognizing a green apple as an "apple"). It is only through a process of feedback that language-learners have their use of categories "corrected" by more experienced language-users. The process of correction may be one-way, as in a teacher-student relationship, or it may be a process of mutual feedback among members of a language community, such as when they work together to refine a coding system to improve their level of inter-rater reliability. Regardless of the language community, the desired end is a high degree of denotative conformity and connotative predictability, and a primary means of reaching that end is feedback aimed at improving a community's understanding of rules of the form "X counts as a Y in context C." Without such shared understanding, the coordinated management of meaning is impossible.

II. Analysis of the Definitions provided in the *Guide to Termination Record Forms.*

[redacted] collects information about drivers' employment histories in part by soliciting Termination Record Forms. The question I address is whether the definitions used

to explain the codes in the "Work Record" section of the Termination Record Form meet the goal of providing "maximum possible accuracy of the information concerning the individual about whom the report relates," as required by the Fair Credit Reporting Act (15 U.S.C. § 1681e[b]).

My assessment of the relevant definitional practices is informed by reviewing the following materials: The initial and amended complaint, copies of depositions (and supporting materials) involving [redacted] and [redacted] a copy of [redacted] "Master User Guide," a document titled [redacted] "Guide to the Termination Record Form," affidavits of [redacted] from the cases of [redacted] and [redacted] affidavits of [redacted] from the cases of [redacted] sample Termination Record Forms, the text of [redacted], various compilations of statistics regarding work history forms, and a copy of the FCRA and relevant regulations.

The focus of this section is [redacted] *Guide to the Termination Record Form* (hereafter GTRF) because this guide "includes the definitions of codes and terms used in the current version of [redacted] Termination Record Form for CDL drivers" (p. 1). This is the only document I found that explicitly attempts to define the key terms used in the Termination Record Form; indeed, the Guide encourages readers to "Use this guide to interpret any term in which [*sic*] you are unsure of the meaning" (p. 1). Plaintiff identifies seventeen phrases or categories that are problematic; I examine each in turn.

"Company Policy Violation." This phrase is defined as code 935 in the GTRF in the following manner: "Driver violated company policies and/or procedures. Use this code only if the other selections in this section do not indicate the company policy violated." It is worth

noting that this "explanation" of code 935 is not a definition in the traditional sense of the word. It is a classic example of a *circular definition*--one that assumes a prior understanding of the term or phrase being defined. It simply repeats the phrase and then provides instruction on when *not* to use the code. It is not an Aristotelian definition, which would require an explanation in the form "A company policy violation is [a kind of] *class name* that has such-and-such *attributes*." There are insufficient criteria provided to infer a clear definitional rule: X counts as Y (a company policy violation) in context C. There is no way to operationalize the phrase except in the crudest fashion, since to qualify for code 935 requires merely one violation of one company policy *or* "procedure."

Apart from lack of definition, understanding the meaning of the phrase "company policy violation" is problematic on several levels. First, no clear exemplar is provided, leaving it up to the person hearing the phrase to provide its "sense." That is, the only connotative predictability one can assume is that the phrase is meant to be pejorative. Second, the phrase is *prima facie* vague, and that vagueness is amplified by the definition when it describes a policy violation as when a driver violated company policies *and/or procedures*. By "vague" I mean that one cannot tell from the phrase what sort of policy and/or procedure was violated, and one certainly cannot ascertain the importance or magnitude of the policy and/or procedural violation. In short, one cannot tell what the words are, in fact, *referring to*. This lack of denotative clarity is made worse by the fact it is defined only by what it is not; that is, the GTRF says to "Use this code only if the other selections in this section do not indicate the company policy violated," which means that one can know only what is *not* being referred to, not what *is* denoted.

An analogy may be helpful in understanding just how meaningless the phrase "company policy violation" is. If I were to say that person A "violated one of the Ten Commandments,"

you would not know what person A did—only that A's action violated one or another commandment. You would not know if person A did something as serious as killing someone, or took the Lord's name in vain, or worked on a Sunday, or coveted a neighbor's car. Because companies have different policies and/or procedures, and religions have different beliefs and norms, a better analogy would be a statement of the form "religious policy and/or procedure violated," which covers everything from mass murder to eating oysters to failing to cross oneself properly. The analogy is useful because *within* various religions, not all sins are treated as equal. Judaism distinguishes among three levels of sin: intentional sin, sins of uncontrollable feelings, and unintentional sins. To state that "someone sinned" does not identify the important attributes of the category—severity and magnitude. Similarly, to state that a company policy and/or procedure was violated does not tell us anything about the severity, magnitude, or type of policy and/or procedural violation that took place. It is, in a practical sense, meaningless. A more useful category system would provide a means to identify the type of company policy and/or procedure violated, as well as the number and magnitude of the violation(s).

"Unsatisfactory Safety Record." This phrase is defined as code 938 in the GTRF in the following manner: "Driver did not meet company safety standards." This is not a circular definition; in fact, it is a sort of operational definition that can be formulated as "A driver has an 'unsatisfactory safety record' when the driver did not meet company safety standards." Unfortunately, the only defining attribute identified ("did not meet company safety standards") is as vague as the previous phrase analyzed, "company policy violation."

Once again, there are insufficient criteria provided to infer a clear definitional rule: X counts as Y (a company safety standard) in context C. Once again, no clear exemplar is provided, leaving it up to the person hearing the phrase to provide its "sense." That is, the only

connotative predictability one can assume is that the phrase is meant to be pejorative. The code and definition in combination are denotatively meaningless because one cannot tell what the words are, in fact, *referring to*. To state that a driver did not meet company safety standards does not tell us anything about the number, importance, or type(s) of standards, nor does it tell us by how *much* a driver did not meet one or more standard. In short, the "definition" provided of "unsatisfactory safety record" renders the code without meaning.

"Excessive Complaints." This phrase is defined as code 912 in the GTRF in the following manner: "An excessive number of complaints have been received regarding the driver's service and/or safety." This is another circular definition, since the "definition" basically restates the phrase being defined and assumes a prior understanding of the phrase.

The definition provided is not an Aristotelian definition, which would require one to identify a set of definitive attributes. Indeed, it is not clear *who* made the complaints, *how many*, *what* the complaints were about, or whether the complaints were *justified*. There are no criteria provided to infer a clear definitional rule: X counts as Y (excessive complaints) in context C. There is no way to operationalize the phrase except in the crudest fashion, since to qualify for code 912 requires merely more than one complaint.

Apart from lack of definition, understanding the meaning of the phrase "excessive complaints" is problematic on two levels. First, no clear exemplar is provided, leaving it up to the person hearing the phrase to provide its "sense." The only connotative predictability one can assume is that the phrase is meant to be pejorative. Second, the phrase is denotatively vague--one cannot tell what the words are, in fact, *referring to*.

"Cargo Loss" and **"Equipment Loss."** These phrases are defined in the GTRF as codes 913 and 917, respectively, in the following manner: "Cargo" or "equipment" "was lost, stolen,

damaged or destroyed while assigned to or under direct responsibility of driver." The problem with these definitions is somewhat different than the previous phrases and definitions. In these cases, enough of a definition is provided that one can formulate a linguistic rule of the form "*cargo/equipment loss* occurs when cargo/equipment is lost, stolen, damaged, or destroyed in a particular context; namely, when assigned to or under direct responsibility of the driver."

The problem is not so much one of denotative vagueness as it is an ambiguous overabundance of possible specific referents. The phrase "cargo loss" could refer to events as disparate as having one's cargo stolen, swept away in a flood, damaged by lightning, or destroyed by vandals. The problem is that a reader of such a report must guess which sort of loss occurred, how serious it was, and who (or what) was the cause.

Given that the *purpose* of the employment history records provided by [redacted] is to aid employers in making hiring decisions, one must evaluate the suitability of the definitions in light of that purpose. That is, do the definitions of the categories identify the attributes important for potential employers? In these cases, they do not, for the simple reason that the definitions do not make clear whether the cargo or equipment loss was *significant* or whether the loss was the driver's *fault*. A more useful category system would provide a means to: A) indicate whether the cargo or equipment was lost, stolen, or damaged, B) estimate the value of the loss, and C) attribute responsibility for the loss. Or, if there is only space for one category, it would be operationalized in such a way to make the information more useful, such as "cargo loss valued in excess of $500 due to driver malfeasance."

"Quit/Dismissed During Training/Orientation/Probation." This phrase is explained as code 933 in the GTRF in the following manner: "Driver did not complete company training, orientation and/or probation. If the driver quit or was dismissed during orientation, leave

sections 12, 13 & 14 blank and do not provide further information to section 15." The second sentence is not a definition, since it is only an instruction as to how to complete other portions of the Termination Record Form. The first sentence is again a classic example of a circular definition that does nothing more than repeat the category label.

Again the problem is that the label has an ambiguous overabundance of potential referents that makes the code unrevealing (meaningless) with respect to identifying driver attributes. It is not clear *when* in the employment process the event occurred, *who* initiated it, or *why*. A reader of such a report must guess, and the range of possibilities is so broad that one cannot make any confident inferences about a driver. Despite this lack of denotative clarity, it is obvious that whatever connotative meaning the label has is negative. "Quit" attributes the cause of the termination event to the driver in pejorative manner. "Dismissed" attributes the cause of the termination event to the employer, again in a manner that is derogatory to the driver.

It would not be difficult to restructure this category to make it more denotatively meaningful and less connotatively negative by indicating *when* the termination event occurred (including whether it was pre-contractual), *who* terminated the relationship (driver or employer), and providing a check-off list of the most common *reasons* for such termination.

"Eligible for Rehire: No." This phrase is explained as code 003 in the GTRF in the following manner: "Driver is ineligible for rehire based on current company standards." This explanation is another example of a circular definition that does nothing more than repeat the category label. The only attribute clearly denoted is that the driver is not eligible to be rehired (which clearly carries a negative connotation); however, the *rationale* for such ineligibility collapses back into one of the vaguest expressions found in the Termination Record Form—"based on current company standards." Again, a reader has no idea what company standards

have informed a decision that the driver is not eligible for rehire, and thus the reader learns nothing about the particular attributes of the driver. Though checking this code makes it clear what the driver's status is with respect to the company completing the form, it conveys no useful information about the driver's abilities. Beyond that company-specific rehiring status, the category is denotatively meaningless.

"Other." This phrase is defined as code 999 in the GTRF in the following manner: "Other: Anything other than items listed above (see 199)." Code 199 says "Other: Anything other than items listed above. This space is provided for your documentation. [redacted] will record 'other' only." Obviously, this category is denotatively meaningless and the category is not defined in any positive sense. There is no way to know what the category is referring *to*, only what it is *not*.

Categories identified as "other" are generally unhelpful in coding schemes. Consider the following example: Let us say that a department store wants to track the reasons that customers return articles of clothing that were purchased at that store. A set of categories might include "wrong size," "garment flawed," or "gift return" and such information could assist both the customers and the store to improve its future service. An unexplained "other" category would be useless because it does not *refer* to anything denotatively. It would be completely useless in helping the store understand why merchandise is being returned, since all anyone could infer is that "something" was wrong.

This case is similar. Since the work record is not described as "satisfactory," there is a vague connotative meaning that is negative—"something" was wrong. But no one receiving such information—either the driver or possible employers—would know *what* was wrong, which makes the information functionally useless.

"Personal contact requested." This phrase is defined as code 944 in the GTRF in the following manner: "Company issuing record has further information to provide regarding the driver or for the driver." This is not a typical category code because it does not even attempt to convey explicit information about a driver's performance. Rather, it is a request for action: For unstated reasons, the company issuing the TRF wishes contact with a potential employer *or* with the driver. Because this category conveys no explicit denotative meaning about the driver's performance, it is not clear to me why it belongs in a section labeled "Work Record."

Since it is a category different from reporting a "satisfactory" (code 901), "superior" (903) or "outstanding" (903) work record, there is a vague negative connotation here that there were problems of some sort warranting a personal contact for explanation. Such meaning is vague and indeterminate, however, since the code explanation includes the possibility that the issuing company wishes to contact the *driver* rather than a potential employer.

Other Descriptive Categories. There are eight additional categories that warrant a different sort of evaluation than the phrases and definitions analyzed so far. These categories are provided with a definition in the GTRF, so they are, in a sense, more meaningful than the circular and vague definitions identified previously. However, these categories are still seriously flawed.

Code 924 "**Late Pick Up/Delivery**" is defined as "Failed to make pickup or delivery according to schedule."

Code 926 "**Log Violation**" is defined as "Violation of Federal Motor Carrier Safety Regulations, 'Hours of Service,' part 395."

Code 928 "**No Show**" is defined as "Driver failed to appear on job site without notification or approval of supervisor. Driver has hauled previous loads for the company."

Code 929 "**Failed to Report Accident**" is defined as "Driver violated accident reporting requirements while in the service of the company."

Code 931 "**Quit Under Dispatch**" is defined as "Driver was available for work, assigned a load but quit before load was secured. Driver did not possess a load."

Code 957 "**Unauthorized Equipment Use**" is defined as "Deviated from route or used equipment for purposes not specified by company. (Not intended to be used when the driver has resigned/quit or terminated lease and returned equipment to the nearest company terminal or a location authorized by the company.)"

Code 959 "**Unauthorized Passenger**" is defined as "Passenger in company vehicle contrary to company policy or did not meet company policy requirements covering authorized passenger."

Code 961 "**Unauthorized Use of Company Funds**" is defined as "Driver used company funds for purposes not authorized by company."

There are three major problems with this set of categories. First, though each code denotes some sort of behavior or event, the category name or phrase is sufficiently broad that it is impossible to determine accurately the significance or importance of the violation. The categories do not allow the person completing the form to indicate the magnitude of the offense, its frequency, duration, or severity. For example, code 924 ("late pick up") could be checked whether the driver was 5 minutes behind schedule or 5 days. With respect to all eight categories, there is simply no way to distinguish between events that may be trivial, accidental, or due to factors beyond the driver's control, versus events that might be quite significant, intentional, and due to driver malfeasance.

Second, it is important to note how the categories are open to abuse due to the fact that all eight categories have distinctly negative connotations. The problem is that there could be two cases that are dramatically different (say, for example one driver is 5 minutes late versus another driver who is 5 days late). The negative connotations and harm to the driver's reputation would be identical since, in both cases, the only message communicated is a checkmark in a particular code box. Thus, even for cases that are denotatively quite different, the categories carry equally weighted negative connotations.

Third, drivers are not provided with these definitions, thus for them these categories are practically meaningless. Note that in some cases the definition is subjective and relies on ordinary language use (such as "late pick up" or "no show"), while others have fairly specific definitions (such as "log violation") that refer to specific policies or regulations. In one case a Federal regulation is referenced, while in several others, "company policy" is referenced. Cumulatively, the eight definitions put drivers between a rock and a hard place. On one side are highly technical definitions that drivers are not provided. On the other are vague or circular definitions that are open to anyone's interpretation. In both types of cases, drivers are disempowered from the relevant language community. Neither drivers nor potential employers are put in a situation to determine the accuracy of the report.

The "bottom line" problem with these categories is that there is no opportunity to provide the sort of details or narrative that would allow someone reading the report to produce an accurate interpretation of the events. We use categories to simplify our understanding of a complex world. However, there is a tradeoff between the *scope* and *precision* of categories: The broader and more abstract a category, the greater the range of events that can be described by it. However, what we gain in scope we lose in precision and accuracy, since a broad category will

lump events together that may be quite different. For example, if we only categorized movies into "comedy" and "serious drama," we would have two categories that have a powerful scope, but at a cost of lumping together films that are quite different. *To maximize accuracy*, one would need to subdivide categories more precisely, so we can distinguish (say) between *To Kill a Mockingbird* and *Star Wars* instead of lumping them together.

One of the best ways to understand the deficiencies of the current category definitions is to imagine how they could be improved. In every case, one can easily imagine how additional descriptors and an opportunity to provide a narrative would increase the meaningfulness of the work record. An example of such an improvement is how the category "Quit Under Load/Abandonment" has been elaborated. At one point, code 909 was "Abandonment" and code 937 was "Quit Under Load." I suspect for the very sorts of reasons discussed throughout this report, these categories were reformulated such that there are now *seven* categories covering a range of events instead of only two. This change nicely illustrates my point about the tradeoff between scope and precision. By elaborating the category, one must give up the simplicity of having only one or two categories, but with seven categories one gains precision and accuracy. I have no doubt that all of the TRF categories could be improved in a similar manner.

III. Overall Assessment and Conclusion

The question I address is whether the definitional practices employed by [redacted] in the "work record" portion of the TRF accomplish their stated ends or not. The [redacted] website description of their Employment History File product claims:

- Members receive more complete information in an efficient manner. Reports include information such as reason for leaving, equipment operated, eligibility for re-hire, status, driver's experience, and number of accidents.

- Employers release and obtain objective, factual information without risk. [redacted] Employment History File protects employers from liability because termination records are submitted using a standard, multiple-choice termination form. Non-subjective, industry standard terminology is used to eliminate the possibility of information being misconstrued.
[redacted]

In contrast, Plaintiff contends that the Termination Record Form that [redacted] pays carriers to fill out and return relies on terms that are "vague, ambiguous, incomplete, uncommonly defined, and inaccurate."

My overall assessment is that the Termination Record Forms fall far short of providing "complete information" about a driver's performance. The multiple-choice format does not produce "non-subjective" terminology that eliminates "the possibility of information being misconstrued," as [redacted] claims. Most of the definitional language is so vague or ambiguous that it virtually guarantees that report writers and readers will systematically misconstrue the denotative meanings of the codes. The system of categories as defined in the TRFs does not meet the requirement to "assure maximum possible accuracy of the information." Indeed, in most cases it is difficult if not impossible to ascertain the specific actions, behaviors, and events that the categories are supposed to refer to. The TRF as currently designed is a source of systemic *inaccuracy* in terms of denotative conformity and connotative predictability.

The problem can be diagnosed by returning to a distinction made in section I between *encoding* and *decoding*. The contested categories of the TRF have been designed in an excessively open-ended fashion from the standpoint of encoding. For example, an incredibly broad array of events can be encoded as "cargo loss." From the standpoint of a former employer completing a TRF, it does not matter if a tornado blew away the cargo or if the truck was robbed. No matter what happened or who was responsible or how much cargo was lost, it would all be

encoded by checking box 913. Those who write the reports are given almost no guidance as to how to "encode" specific events or attributes. Without such guidance, errors of overextension (applying a code too broadly or "false positives") and underextension (not applying a code when one should, or "false negatives") are inevitable.

No coding scheme or system of classification is neutral: All guide our attention in particular ways by providing semantic maps for making sense of our experiences, Such maps tell us what is important to notice and what can be neglected, and what is valuable and what is not worth our attention. By keeping the TRF codes to a minimum, the categories are "defined" so flexibly as to make them largely meaningless. **Furthermore, the "flexibility" of the encoding process is what makes accurate *decoding* impossible.** The data have become meaningless because it is impossible for a writers and readers to know what the codes are referring to (denotative meaning) and only vaguely how the codes are evaluating the driver (connotative meaning). The varying frequency of usage of the various codes by different carriers underscores this point—the TRF does not constrain coders, it gives them excessive latitude such that decoders have no clear idea what is or is not being reported.

The data gathered through such a coding scheme does not serve its purpose in assisting employers make informed hiring decisions based on accurate and precise information, and it obviously does not serve the interests of drivers either. Indeed, insufficient information is provided to allow drivers to know what behavior resulted in what sort of evaluation, which makes it extremely difficult to check or dispute the accuracy of such records.

Just how far short of the requirement to "assure maximum possible accuracy of the information" the TRFs are can be seen most clearly by considering how the categories *could have been* constructed and defined in such a way to avoid the problems identified in section II.

In each and every case, the problems identified above could be solved by providing additional codes to distinguish more precise subcategories or by defining the current codes with greater denotative precision.

Since the publication of Claude E. Shannon's "A Mathematical Theory of Communication" in 1948 it has been understood that "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (p. 379). **The question is not only do the people completing a TRF know which box to check, the question is also whether those who subsequently *read* the output can accurately decode the meaning of such checked boxes.** If the message "received" or "interpreted" by the reader of a TRF is significantly different than the message "sent" by the original source, then we have, as it is put in *Cool Hand Luke*, "failure to communicate." That failure can be summarized as a profound lack of clarity and specificity in the "definitions" of the codes, which results in a lack of denotative conformity and connotative predictability.

To press the "diagnosis" a step further, I would suggest that the problems identified stem, in part, to a lack of shared purpose among drivers and carriers in creating the definitions and codes. The TRF is apparently designed wholly to serve the interests of carriers, who are consistently referred to as the "customers" in the depositions of [redacted] The code categories appear to have been defined to minimize the difficulty of filling out the form, while maximizing the power of the carriers over drivers.

If the categories are defined entirely from the carriers' perspective, then drivers are excluded from being what were described in section I as "appropriate authorities." In [redacted] second affidavit in [redacted] she states that "Drivers are not users of our employment history reports" (§24). Despite the claim that [redacted] uses definitions that follow industry

practice," she acknowledges that "I routinely encounter drivers who dispute their employment history reports because they do not understand the meaning of the terms in the report" (§24). The power to define is entirely in the hands of [redacted] Services, which apparently does not include any driver representatives on the [redacted] Advisory Board.

Furthermore, there does not seem to be any formal or institutionalized process of providing *feedback* that assures drivers a role in refining the code definitions. Drivers are not provided a copy of the "Guide to Termination Record Form," which includes the list of "definitions" I analyzed above. This lack of information obviously hampers drivers' ability to understand how or why their work record has been evaluated in a particular manner, and it makes the task of disputing a particular evaluation extremely difficult. Any informal or formal means of dispute resolution is hampered. Furthermore, the vagueness and ambiguity of the language function strategically to deflect responsibility by maintaining a kind of "plausible deniability" (Walton, 1996) about the "meaning" of TRF codes. That is, the vaguely negative connotations of the categories discussed above create a negative "presumption" about a driver, but because the form stops short of providing clear denotative meanings, [redacted] can deny specific inferences made from ambiguous codes. The TRF thus functions as a form of systematic "innuendo" about drivers and [redacted] avoids assuming a reasonable "burden of proof" for what is inferred from the vague categories (cf. Walton, 1996).

Giving carriers "definitional hegemony," or near-total authority over how a driver's history is encoded, functions to infantilize drivers in the language community that makes up the trucking industry. By denying appropriate authority or adequate opportunity for feedback for drivers, the category codes are potentially open to a good deal of abuse. The definitions of the codes are so vague, ambiguous, and/or circular that they can be stretched to describe just about

anything. Whether they have been abused is a question I am not in a position to answer, but I can say with confidence that the codes are very poorly designed and open to abuse.

To conclude: The definitional practices as found in the *GTRF* and in the various documents I studied associated with this case fail to provide "maximum possible accuracy of the information concerning the individual about whom the report relates." The TRF is not designed to provide accurate denotative or connotative meaning in terms of driver attributes. The category codes are vague or ambiguous—they do not provide sufficient guidance to promote either denotative conformity or connotative predictability. Because there is a lack of explicit and shared definitional purposes, and because drivers are not treated as appropriate authorities or provided an institutionalized opportunity for feedback, the definitional practices are seriously flawed.

Works Cited

Bowerman, Melissa. 1976. "Semantic Factors in the Acquisition of Rules for Word Use and Sentence Construction." In Donald M. Morehead and Ann E. Morehead, eds., *Normal and Deficient Child Language*. Baltimore: U. Park P. 99-179.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: M.I.T. P.

Frege, Gottlob. 1892. "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50. Translated as 'On Sense and Reference' by M. Black in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds. and trans.), Oxford: Blackwell, third edition, 1980.

Matthews, Alexander. 1998. *A Diagram of Definition: The Defining of Definition.* Assen, The Netherlands: Van Gorcum.

Rey, Alain. 2000. "Defining Definition." In *Essays on Definition*, J. C. Sager (ed.). Amsterdam: John Benjamins Publishing.

Roberts, Kenneth, and Frances Degen Horowitz. 1986. "Basic Level Categorization in Seven- and Nine-Month-Old Infants." *Journal of Child Language* 13: 191-208.

Robinson, Richard. 1950. *Definition.* Oxford: Clarendon P.

Schiappa, Edward. 2003. *Defining Reality: Definitions and the Politics of Meaning.* Carbondale: Southern Illinois U. P.

Schramm, Wilbur (Ed.). 1954. *The Process and Effects of Mass Communication*. Urbana: U. of Illinois P.

Senft, Gunter, ed. 2000. *Systems of Nominal Classification*. Cambridge: Cambridge U.P.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27: 379-423, 623-656.

Walton, Douglas. 1996. "Plausible Deniability and Evasion of Burden of Proof." *Argumentation* 10: 47-58.

Report on Validity Issues in *[redacted]*

Glenn Beamer, Rutgers University
July 7, 2005

Report on Validity and Reliability Issues in *[redacted]*

Glenn Beamer, Rutgers University
July 7, 2005

Executive Summary

1. Survey and reporting forms should provide valid and reliable information that reflects past events and respondent characteristics. In this case validity has two components – content validity and measurement validity. Content validity is the extent to which individual queries measure characteristics and events such that they can be readily related to a relevant dimension of an employment history. Measurement validity is the degree to which individual queries accurately, systematically, and dependably report the extent of characteristics and frequency of events. Reliability is the extent to which individual queries are likely to be reported accurately by numerous recorders taking into account human error and perceptual differences that can contribute to varied recordations. By maximizing the extent to which any two recorders will accurately report the same events and characteristics, regardless of other factors external to the survey, researchers minimize the unreliability of survey data.

2. Content validity requires that individual queries accurately measure individual events and characteristics related to a specific dimension of an employee's character or work history. With respect to employment histories, for a survey to have construct validity questions about individual events and characteristics should be connected to underlying dimensions of employment history such as safety, customer relations, trustworthiness, timeliness, and policy compliance. The flawed design of the [redacted] Termination Record form, its lack of topical organizations, and the lack of connections among topics and individual measures eliminate clarity from the termination record form work record section and render it invalid with respect to its contents.

3. Measurement validity is the extent to which individual queries accurately portray the frequency of an event or the extent of a characteristic. In the [redacted] Termination Record form work record section the biased and inadequate scale used for 17 out of 20 individual measures as well as ambiguous instructions and definitions render the measurements invalid.

4. Reliability is the extent to which recorders succeed or fail at accurately representing, literally re-presenting, an employee's character and work events. Whereas measurement validity pertains to how well or poorly individual queries reflect event frequency or the extent of character traits, reliability results from respondents consistently and accurately portraying events and characteristics. Unreliable data can result from human error, from recorders having inconsistent understandings of how and when to record events and characteristics (i.e. ambiguous or uninformative survey directions), and from recorders inconsistently using different categorical responses to reflect the same event or characteristic (i.e. inconsistent application of the survey directions). The analysis of the data from a sample of 5000 [redacted] Termination Record forms and from work record code user profiles of 3603 carriers reveals that driver characteristics and employment events are inconsistently and inaccurately reported and that [redacted]

Termination Record forms fail to provide reliable representations of drivers' characteristics and work histories.

5. The definitions and directions provided by *[redacted]* for recorders using the [redacted] Termination Record form provide inadequate guidance and contribute to the unreliability of the data produced by carriers.

6. Data generated by [redacted] Termination Record forms suffers from the lack of content validity, invalid measurements, and unreliable reporting. Consequently:

    a. *[redacted]* and carriers using the [redacted] Termination Record form have no means by which to organize, aggregate, or index data such that they can make meaningful inferences about an individual driver's professional characteristics or event history (i.e. the [redacted] Termination Record form lacks content validity).

    b. Carriers using [redacted] Termination Record forms can have no confidence that they are assessing an employee's work history as it pertains to the frequency of events or the extent of characteristics (i.e. the [redacted] Termination Record form lacks measurement validity).

    c. Carriers using the [redacted] Termination Record form can have no confidence that recorders have consistently and accurately reported drivers' characteristics and employment histories (i.e. the data produced by [redacted] Termination Record forms are unreliable).

7. Statistical analyses of [redacted] Termination Record forms reveal that hundreds of carriers use individual codes in ways that are statistically deviant from ranges expected in varied samples and that indicate systemic distortions. Carriers used codes at rates 4 to 33 times greater than would have been expected in the absence of systemic distortions.

8. A reliability analysis of [redacted] Termination Record form work record codes produces a reliability coefficient of only 0.157 on a 0 to 1.0 scale. Reliable data typically have corresponding coefficients in excess of 0.7 if not 0.8. The estimated error rate for [redacted] Termination Record forms is 60%.

Introduction: Content Validity, Measurement Validity, and Reliability

Content validity, measurement validity, and reliability are three qualities of survey research relevant to the case of *[redacted]*. The survey is the [redacted] Termination Record form and the measures are the individual queries designed to record events and their frequency and driver characteristics and their extent.

This report begins with a section in which I define content validity and measurement validity. I explain content validity, measurement validity, and their pertinence to the [redacted] Termination Record form. Section 2 provides an analysis of the content and measurement validity of the [redacted] Termination Record form's work record queries. I rely upon the termination record form and its instructions to carriers regarding how to record events and driver characteristics. In section 3 I evaluate the termination record form based upon the data produced by carriers who have used the [redacted] Termination Record form. In other words section 2 presents an *ex ante* evaluation of the form as it is presented to carriers and section 3 presents an *ex post* evaluation in which I assess employment histories produced by carriers using the [redacted] Termination Record form.

The evaluation in section 2 establishes that *[redacted]* fails to have content validity in the work record section of the termination record form and that the individual work record measures are inadequate and biased. Because the measures' scales are inadequate and biased the measures themselves are invalid. The analysis in section 3 reveals that carriers behave aberrantly when using work record codes and that carriers do not share a common understanding of how and when to use different codes.

Section 4 presents a basic analysis of the termination record form work record section and a statistical analysis focusing on the reliability of the employment histories produced by the work record section. The basic analysis reveals that at least ten percent of termination record forms are incomplete and cannot be relied upon as meaningful work histories. The statistical analysis reveals that carriers use the work record section inconsistently and the data produced to represent employment events and driver characteristics are unreliable. I estimate an error rate for the work record section of 60 percent.

Section 1: Content Validity and Measurement Validity

Content validity and measurement validity are qualities of empirical reports such as employment histories produced from survey forms such as employee evaluations. Content validity refers to how clearly individual survey measures relate to topics relevant to the research subject. Measurement validity regards the extent to which survey queries accurately gauge and denote individual items relevant to various topics. Content validity and measurement validity are not absolutes (Carmines and Zeller, 1978; Litwin, 1995). Research instruments are evaluated as having more or less content validity and more or less measurement validity with respect to the individual measures, topical categories, and overall subject (Babbie, 2004). In *[redacted]*, the research instrument is the [redacted] Termination Record form and the data it produces are drivers' employment histories. The [redacted] Termination Record form's subject is the employment history of an individual driver. Categorical topics within an employment history could include safety, professional standards and conduct, and customer relations (U.S. Department of Labor, 2005). Individual queries are the measures that comprise a driver's profile for each topic and these topics then comprise an employment history.

Evaluating content validity is a qualitative exercise that is nevertheless rigorous and thorough (Babbie, 1994). By examining the organization and evaluating the clarity of a survey *Termination Record form* such as the termination record form, one can assess its content validity. Both content and measurement validity can be evaluated by comparison with similar surveys (Litwin, 1995; Fowler, 1995)). Attached as Appendix A by way of illustration are the termination record form, two driver evaluation forms used from public school districts. Attached as Appendix B by way of illustration is an alternative work record section of the termination record form that I composed. I compare the termination record form to these employment history reports. These comparisons lead to specific changes that if made would increase substantially the *[redacted]* Termination Record form's validity.

Evaluating Content Validity and Measurement Validity. There are three criteria that we can use to assess the content and measurement validity – clarity, scale, and bias (Babbie, 2004; Judd, Kidder, and Smith, 1991; Litwin, 1995; Carmines and Zeller, 1978). Clarity relates directly to content validity (Babbie, 2004). Scale and bias relate to measurement validity (Litwin, 2004; Fowler, 1995). I discuss each of these in turn.

Content Validity & Clarity. In order for surveys to be meaningful, the individual measures and topics researched should be clearly identified for both the research producer and

consumer (Babbie, 2004). In *[redacted]*, the research producers are previous employers who submit [redacted] Termination Record forms to *[redacted]*. *[redacted]* in turn uses submitted termination record forms to produce employment histories. The research consumers are prospective employers who evaluate the employment histories. Clear definitions and categorical organization increase the content validity of surveys by enhancing their straightforward production and interpretation (Babbie, 2004).

Content validity involves identifying and using appropriate measures as components of topics that are relevant to the research subject (Babbie, 2004; Litwin, 1995). The [redacted] Termination Record form is an instrument that carriers use and which *[redacted]* collects to produce employment histories. As a research instrument the [redacted] Termination Record form should include measures of dimensions of truck drivers' employment histories. Thus the subject of the research being conducted by a prospective employer is the driver's employment history. Three topics considered relevant to that research subject, the employment history, would likely include driver safety practices, driver professional conduct, and driver customer relations. Individual measures of driver safety could include measures of events such as accidents, measures of compliance with safety regulations posted by the company and law enforcement, and measures of third party observations such as merit citations or records of complaints from drivers or coworkers. Measures of professional conduct could include attendance and timeliness, schedule maintenance, relations with coworkers, relations with supervisors, and communication records.

The first step toward establishing clarity is to identify the relevant topics and the measures related to those topics for producers and consumers. Typically this step involves organizing measures in groups that correspond with a topic. In situations in which a measure could apply to more than one topic, clear organization can direct information producers and consumers as to which topic or topics the measure is intended to relate (Litwin, 1995).

A second step in establishing content validity is to organize a survey by categorical topics (Fowler, 1995). For employment histories topical organization should enhance both content and measurement validity by focusing recorders on an employee's experiences and characteristics for each area of investigation.

To understand content validity an example is helpful. Suppose we have a classroom with twenty children and we are interested in their size. Size is a broad term – you might think of weight and I might think of height. Both height and weight are valid components of child's

size. Just as safety and professional conduct are valid topics for an employment history, size and weight are valid topics if our research subject is "childrens' size." If a third person came to our classroom with an eye chart and stated that she was going to test children's eyesight, we would understand that eyesight or visual impairment is not a topic related to size. Eyesight is a dimension of the subject "sensory ability," along with feeling, hearing, and taste. Height and weight are clearly dimensions of size and can be assessed as having content validity with respect to the subject size. Eyesight is an invalid dimension of size but a valid dimension of "senses."

Content validity also refers to the extent to which measures comprehensively cover a dimension of the research subject (Babbie, 2004, pp. 143-146). Just as height is only one dimension of size, an employment history that focused only on schedule maintenance would be lacking. In order to have an employment history with content validity other dimensions such as safety and customer relations should be included.

Measurement Validity, Scale, and Bias. Measurement validity refers to the extent to which an instrument or query accurately and fairly represents the frequency of an event or the extent of a characteristic (Fowler, 1995; Judd, Kidder, and Smith, 1991). Valid measures reflect the range of possible outcomes for queries relevant to a topic (Babbie, 2004; Carmines and Zeller, 1978). Valid measures should provide sufficient information such that research consumers, in *[redacted]* potential employers, understand that the information accurately reflects the extent of characteristics, such as children's heights, and the frequency of events.

Standard scales enhance measurement validity and help ensure that recorders understand and can use a range of available responses that correspond to the range of events and characteristics (Fowler, 1995). Survey measure scales should have two characteristics – they should be adequate and they should be consistent (Fowler, 1995). Adequate scales are appropriate for measuring the size, extent, and frequency of events and characteristics. This means that children's weight measurements would be made in pounds and ounces and not tons. Freight may be appropriately measured in tons and not ounces or milligrams.

When categorical responses, such as not satisfactory, satisfactory, and excellent are used in measurements, recorders should be given directions about the thresholds for using the different ordinal categories (Fowler, 1995; Litwin, 1995). Thresholds and guidelines for recording categorical responses enhance the consistency with which researchers use measurement scales. For example a researcher interested in employees' safety habits might

direct recorders to identify unsafe employees as those having "two or more accidents resulting in physical injury to themselves or others within the last year." This direction has three important components: 1) an empirical threshold of two or more accidents, 2) consequences of the accident which are physical injuries to the employee or others, and 3) a time frame that reflect the period for evaluation – the last year (Fowler, 1995).

In addition to providing responses for the range of outcomes adequate scales may reflect the context in which a given measure is taken. If we are measuring the frequency of school absences, we may want some indication of the number of days missed due to excused illnesses compared to unexplained absence. If we are measuring corporate performance, we may want to measure profits from operations and profits from asset sales. Conflating these two profit measures might provide a valid measure of "total profitability," but it would be incomplete and potentially misleading if we were looking to invest in the company.

In assessing adequacy we should ask "Are the measures adequate for the task at hand?" If we are designing a playhouse for a classroom full of children then knowing their heights and weights is adequate to build a structure in which the children won't get stuck, hit their heads, or fall through the platform. If we are sewing clothing for our classroom of children, then knowing their height and weight is probably not adequate for making each student a shirt. To sew shirts we would want to know the children's neck circumferences, their shoulder widths and their sleeve lengths. Adequate measures reflect the range of outcomes including the context in which events occurred or characteristics have been exhibited.

Unbiased. Good measures do not motivate recorders to classify subjects inappropriately. In addition to providing sufficient categories (adequacy) measures should not be subjective or should not focus on only negative characteristics and events or only on positive characteristics and events. Whereas comprehensive response categories and appropriate units of measurement contribute to adequacy, balanced identification and question wording create unbiased measures (Fowler, 1995).

Unbiased measures identify neither negative nor positive events and characteristics exclusively. Unbiased questions identify and record different types of events and characteristics and the frequency with which they occurred or the extent to which they were evident during employment histories. Unbiased questions should allow research consumers to discern, without great work, the extent of a characteristic relative to other survey respondents or the relative frequency of events. Biased questions often eliminate large subsets of respondents (in this

case drivers) from measurement because the questions do not apply or provide for inadequate responses that allow only negative inferences. Unbiased questions and responses provide the research consumer with information that can lead to either positive or negative inferences (Fowler, 1995).

To better understand measurement validity consider again a classroom of 20 children ages six to ten and for whom we want to record their heights. If a researcher came to the classroom with a bathroom scale, she could not record students' heights; she could record their weights. The bathroom scale is an invalid measurement instrument for recording height. Rulers, yardsticks, and a tape measure would all be valid instruments for measuring height. However, if a researcher brought a yardstick that was 40 inches long and claimed it measured three feet this would not be a valid research instrument. This 40-inch yardstick underreports children's heights by four inches per every three feet and its resulting measures are invalid. A yardstick is appropriate for measuring children's height and its scale, feet and inches, is both appropriate and adequate. But a 40-inch yardstick is a biased measurement instrument and thus yields biased and invalid measurements.

Section 2: Content Validity, Measurement Validity, & the *[redacted]* Form

For the analysis of content validity and measurement validity in the [redacted] Termination Record form I focus first on the form and its instructions. Neither the [redacted] Termination Record form nor its instructions establishes content validity because neither clearly identifies relevant categories of driver experiences and character.

The [redacted] Termination Record form's "work record" section has two categorical areas – work history and load abandonment. The first topic is denoted by work codes 901 through 938 and codes 957 through 999. The instructions include guidance for code 915, which no longer exists on the termination record form. Although *[redacted]* references "categories" for carriers to identify drivers as satisfactory, superior, or outstanding, *[redacted]* does not explicitly identify these categories on the instructions. For drivers to be recorded as satisfactory, superior, or outstanding *[redacted]* instructs employers that such drivers must meet or exceed individual company standards "in all categories." Presumably these are performance categories such as employee relations, customer relations, and safety but *[redacted]* makes no effort to identify the relevant categories nor does *[redacted]* specify which individual measures relate to these unidentified categories. Therefore there is no means by which to establish content validity for the Work Record section of the [redacted] Termination Record form when it is considered as a whole survey instrument. The failure to establish content validity rests squarely with *[redacted]* and *[redacted]*.

For comparative evaluations I performed a "Google" search for "Driver Evaluation Forms." I printed forms from the University of Alabama Annual Employee Performance Evaluation, the Tacoma Public Schools Bus Driver Evaluation, the Rockingham County (Virginia) Public Schools Bus Driver Evaluation, and the State of Idaho (Bus) Driver Evaluation forms. All of these forms have identified work history categories each category has multiple measures. This search, form review, and printing required approximately forty minutes. The Tacoma and Rockingham County forms are reproduced in Appendix A.

The Tacoma and Rockingham driver evaluation forms have two features that provide greater content and measurement validity than the [redacted] Termination Record form work record section. Their content validity is established because the forms are organized into three or four work history categories such as professional conduct and interpersonal relations. The forms then have subordinate measures that clearly relate to these categories. The Tacoma

Public Schools forms organizes driver characteristics and employment events into 3 categories with 26 measures, the Rockingham County Schools form organizes driver characteristics and employment events into four categories with 30 measures. Taken together the topical profiles provide comprehensive reflections of bus driver characteristics and employment events.

The Tacoma and Rockingham County forms' organization provide clear direction for potential employers assessing drivers' work histories. In comparison the [redacted] Termination Record form conflates topical categories of employment history and compounds its disorganization by including identifiers that relate to the context of employment termination and requests for communications among carriers (codes 950 through 956 and code 944 respectively). Although *[redacted]* instructs previous employers to use all identifiers that apply, the vast majority of previous employers use only a single identifier. Table 1 compares the [redacted] Termination Record form with the Bus Driver Evaluation forms.

The Tacoma and Rockingham forms have standardized ordinal responses for each measurement. For each query recorders are asked to rate Rockingham drivers as 1) does not meet (the standard), 2) needs improvement, 3) satisfactory, 4) and outstanding. For each query on the Tacoma form evaluators are asked to rate drivers as 1) needs improvement, 2) meets expectations, 3) exceptional, 4) not observed. These comprehensive response scales permit evaluators to record an evaluation of every driver for every measure. In instances where the evaluator cannot reliably report an employment event or a driver characteristic the Tacoma form provides a response in which the evaluator can indicate his or her lack of observation. The Tacoma and Rockingham forms provide response scales that yield balanced and comparable measurements for all drivers. The [redacted] Termination Record forms provide neither evaluations of drivers for every measure nor balanced response codes that differentiate types of experiences.

Table 2.1: Comparison of Work Record Forms & Employment Histories

Form	Identified Contents	Measures per Category	Format of Measures
Rockingham Public Schools Driver Evaluation Form	1) Organization & Planning 2)Interpersonal Relations & Communication 3) Professional Responsibilities/Qualities 4) Knowledge & performance of job responsibilities	5 4 7 14	Four category ordinal scales: 1) Outstanding, 2) Satisfactory, 3) Needs Improvement 4) Does Not Meet Drivers are evaluated with all 30 measures
Tacoma Public Schools Bus Driver Evaluation Form	1) Job Knowledge & Skills 2) Student Management Skills 3) Personal Qualities	12 4 10	Four category ordinal scale: 1) Exceptional 2) Meets Expectations 3) Needs Improvement 4) Not Observed Drivers are evaluated with all 26 measures.
[redacted] Form/Work Record Section 900	None identified but drivers must meet or exceed carrier expectations "in all categories" to be evaluated as satisfactory, superior, or outstanding.	20 work event and driver character measures plus 7 measures regarding location of termination.	Check if event occurred/characteristic applies. Carriers may circle as many as apply. Drivers are evaluated with zero to eighteen measures.

One potential objection to adopting a form similar to the Rockingham or Tacoma forms could be that these forms utilize an entire page and require employees' supervisors to record to thirty and twenty-six measures. Because of their organization, these forms require less consideration than the [redacted] Termination Record form. Supervisors using the school district forms understand that they record every measure for every driver. Carriers using the [redacted] Termination Record forms must decide which "identifiers" to use and which to

disregard. In cases where "identifiers" are not used, a potential employer cannot infer whether the identifier is unused because it does not apply or whether it is unused because the previous employer chose to ignore or skip the identifier. With the Rockingham and Tacoma forms unrecorded measures reflect incomplete employment histories. With the [redacted] Termination Record form unrecorded measures may or may not reflect incomplete employment histories. This point is relevant to the issue of data reliability as well as measurement validity. The superior measurement validity of the Rockingham and Tacoma forms enhances their reliability. The biased and inadequate scales in the termination record form likely compound their unreliability.

Fourteen of the twenty *[redacted]* work history measures ask previous employers to identify single, idiosyncratic events in a driver's employment history.[1] These measures provide negative impressions about drivers who may have otherwise satisfactory, superior, or outstanding records. In contrast the Rockingham form uses a four category ordinal scale ranging from "Does Not Meet (Standards/Expectation)" to "Outstanding." The Tacoma form uses a three category ordinal scale ranging from "Needs Improvement" to "Exceptional." The Tacoma form also includes a category "Not observed." This latter category enhances both measurement validity (events and characteristics that may not exist are not mistakenly represented) and the form's reliability. The bus driver evaluations use qualifiers such as "As needed" to indicate whether drivers appropriately complied with school district policies. The [redacted] Termination Record form' s lack of such qualifiers deprives the measures of contextual representation and further undermines its measurement validity.

These fourteen measures deprive drivers of comprehensive employment histories. Identifying a driver as "polite and on time one day" would not provide a comprehensive employment history, and it is equally skewed and uninformative to use idiosyncratic negative events to reflect drivers' experiences and characteristics. These measures are equivalent to using a yardstick to measure children's heights but only measuring children who are less than three feet tall. Any children over three feet tall are left unrecorded and thus consumers of the children's height research are left with an extremely biased representation of children's heights.

---

[1] These fourteen identifiers are 913 – Cargo Loss, 915 – Falsified Employment Application, 917 – Equipment Loss, 924 – Late Pick Up/Delivery, 926 – Log Violation, 928 – No Show, 929 – Failed to Report Accident, 931 – Quit Under Dispatch 933 – Quit/Dismissed During Training and/or Orientation, 935 – Company Policy Violation, 957 – Unauthorized Equipment Use, 959 – Unauthorized Passenger, and 961 – Unauthorized Use of Company Funds.

Information consumers know only about relatively short children and have no means of knowing how *relatively* short they are because no data are recorded for taller children.

These same fourteen measures in the [redacted] Termination Record form inquire about only negative events such as "Late pick up," "unauthorized equipment use," or "cargo loss." These measures are biased. Unbiased measures would inquire about drivers' schedule maintenance, equipment stewardship, and cargo responsibilities respectively. Because the measures are biased potential employers cannot know drivers' employment histories and cannot evaluate the frequency of negative events or the extent of negative driver characteristics. Correspondingly these measures provide no response space or response category for drivers who perform tasks well (i.e. have only positive driver characteristics) or who experience no adverse employment events. Along with not providing for positive identifications the *[redacted]* measures fail to provide previous employers with categorical responses indicating whether or not drivers were culpable for negative employment events.

These comparisons establish that the [redacted] Termination Record form could have substantially more content validity by organizing its work history identifiers by categorical topics. The [redacted] Termination Record form could have substantially more measurement validity by providing comprehensive and balanced response codes. With this overall analysis in mind I now turn to an evaluation of the individual measurements in the work history section of the [redacted] Termination Record form.

Evaluation of Individual Measures. The focus of this section is the [redacted] Termination Record form work record measures, coded 901 through 999, and their respective instructions in the [redacted] Termination Record form and in the *[redacted]* Services Guide to the Termination Record Form. Excluding the seven codes, 950 through 956, that deal with load abandonment, there are twenty measures. Among these twenty measures 10 are unclear, 11 are inconsistent, 17 have inadequate responses, and 17 are biased. Additionally one measure, code 944 – personal contact requested, is not a measure of employment history.

In what follows I assess the content validity and measurement validity of the individual termination record form based upon categories of employment history such as safety and professional conduct. The Guide to the [redacted] Termination Record form does not provide specific guidance about these contents and therefore I created my own.

"Satisfactory, Superior, and Outstanding." These measures are intended as summary measures of employment history. *[redacted]* does not organize its form such that there is a

measure for "unsatisfactory employees," but *[redacted]*s instructions reflect that codes 901 through 903 are summary or aggregate identifiers.

[redacted] instructs employers to identify drivers as satisfactory if the driver "meets minimum company standards of performance in all categories", and *[redacted]* instructs employers to identify a driver as superior if the driver "exceeds minimum company standards of performance in all categories." *[redacted]* does not identify these constituent categories and leaves their inclusion or exclusion to carriers' discretion. This lack of definition creates an inconsistent measure akin to including or excluding the length of children's heads or legs for height measurements. Because each carrier may include or exclude any number of categories the definitions for satisfactory and superior are very likely to be inconsistent.

Based on *[redacted]*s instructions, some companies may include only a single category (e.g. driver safety) for an evaluation of satisfactory or superior while other companies may include a dozen or more categories. Because carriers subjectively and variously define individual dimensions of drivers' employment histories the codes *satisfactory* and *superior* lack clarity and content validity. Previous employers are free to define by inclusion and exclusion their topical components in the summary evaluations "satisfactory" and "superior" with no means of identifying those topics for potential employers. Potential employers cannot discern the components that comprise satisfactory and superior drivers. Table 2.2 explains why the codes satisfactory, superior, and outstanding have neither content nor measurement validity.

Table 2.2: Content and Measurement Validity

Work Record Item	Content Validity	Measurement Validity	Explanation
901- Satisfactory	Invalid	Invalid	■ No identification of categories included (inconsistent definitions) ■ Varied thresholds for "minimum company standards" (lack of scale)
902 – Superior	Invalid	Invalid	■ No identification of categories included (inconsistent definitions) ■ Varied Thresholds (lack of scale) ■ Rating may not be attainable (inadequate scale)
903 – Outstanding	Invalid	Invalid	■ No identified categories (Inconsistent definition) ■ Undefined threshold for "Outstanding" (lack of scale)

Satisfactory: Driver meets minimum company standards of performance in all categories. With respect to measurement validity for code 901 satisfactory drivers must meet minimum standards but potential employers have no means of knowing what these minima are. One carrier may have a minimum standard that prohibits any swearing. A driver who said "Hell" would not be coded as satisfactory but rather as "935 company policy violation." The lack of threshold indicators will lead to inconsistent identifications.

Superior: Driver exceeds minimum company standards of performance in all categories. *[redacted]* instructs previous employers to identify drivers as "superior" if they "exceed" company standards. A consideration for this code is that exceeding a standard may be impossible. A company may have a standard that drivers are not to have accidents. A driver who has no accidents has met the company standard, but cannot "exceed" that standard. A driver cannot have a "negative" accident that would in some sense exceed the company standard of not having an accident. Because the measure *superior* may be unattainable it is inadequate and this decreases further its measurement validity.

Outstanding: Driver's performance is outstanding in all categories. This instruction suffers not only from the lack of categorical definition but also from a lack of reference for identifying "outstanding" which is used as both the measure and the definition of the measure. The threshold for outstanding drivers is subject to carriers' varying standards. These varied standards in turn contribute to inconsistent measurement.

Comparing the instructions for recording superior and outstanding evidences their lack of measurement validity. *[redacted]* instructs carriers to code as "superior" only those drivers who exceed their individual company standards for "all categories." "Outstanding" drivers are defined at the discretion of the employer – there is neither an identified threshold nor categorical definition. A carrier using a single category – safety – may code an employee as outstanding because the driver safely delivered a load through a hurricane once. Another carrier may decline to identify a driver who has served dependably for a decade but then was late for his final delivery. Because the termination record form requires that drivers exceed company standards "in all categories" and because this driver can be identified using a measure that relates to a single idiosyncratic event the previous employer may choose not to identify this superior driver as such. With clearly identified employment history topics that in turn constitute identification as satisfactory or superior potential employers could discern drivers who failed to

meet a previous employers' standards in a single category verses drivers' who failed to meet previous employers' standards in numerous categories.

Complaints: An excessive number of complaints have been received regarding the driver's service and/or safety. This measure and its instructions lack content and measurement validity. With respect to content validity, the measure "complaints" conflates reports about safety, customer service, and professional conduct. The [redacted] Termination Record form lacks guidance about relevant sources of complaints, which could include co-workers, customers, law enforcement officials, and the general public. Table 3 details the flaws in the complaint code and its instructions. The lack of a clear categorical relationship deprives the code of content validity. The lack of guidance regarding complaints about safety practices or professional conduct renders the measurement inconsistent.

The complaints code lacks a threshold for the number of complaints required to be "excessive." A potential employer can only infer that the number of complaints is more than one. This lack of an identified threshold deprives the code of measurement validity.

Table 2.3: Customer Relations Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
911 – Complaints	Invalid	Invalid	■ Conflates service and safety related complaints (lack of clarity) ■ Threshold for excessive is undefined and may vary widely across carriers (inconsistent scale) ■ No categorical response available for employees without recorded complaints (Biased measure).
924 – Late Pick Up/Delivery	Valid	Invalid	■ Identifies a single idiosyncratic negative event (biased measure) ■ Measure fails to identify drivers who routinely met scheduled pick ups and deliveries (Biased scale) ■ Measure makes no provision for driver's responsibility (Inadequate scale)

Late Pick Up/Delivery: Failed to make pick up or delivery according to schedule. The code Late Pick Up/Delivery could be readily identified as a constituent of customer relations or professional conduct and thus has content validity. There are three flaws that render code 924's measurement invalid. 1) Late Pick Up/Delivery identifies a single idiosyncratic negative

event. This focus biases the measure such that the driver is represented negatively no matter his or her overall record of schedule maintenance. 2) The measure makes no provision for drivers who met their pick up and delivery schedules and thus excludes most drivers. This inadequacy further detracts from the code's measurement validity. 3) The measure provides no indication of whether or not the driver was responsible for the single idiosyncratic late pick up or delivery. Some previous employers may use this code when drivers could not avoid being late (because of severe weather or unsafe conditions) while other previous employers may use this code only when lateness is clearly the driver's responsibility (e.g. the driver overslept).

Professional Conduct. The next eight codes I evaluate broadly relate to drivers' professional conduct. *[redacted]* makes no effort to identify or define this category. Table 4 presents summary evaluations for codes 913 – cargo loss, 917 – equipment loss, 915 – falsified employment application, 926 – log violation, 935 – company policy violation, 957 – unauthorized equipment use, 959 – unauthorized passenger, and 961 – unauthorized use of company funds.

Table 2.4: Professional Conduct Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
913 – Cargo Loss	Invalid	Invalid	■ Could relate to customer satisfaction, safety, or employee trustworthiness (lack of clarity) ■ No assignment of responsibility for cargo lost (driver, weather/act of God, theft, equipment poorly maintained by carrier) – (Inadequate scale) ■ No response category to reflect drivers who experienced no losses (Biased measure)
917 – Equipment Loss	Invalid	Invalid	■ Could relate to safety, trustworthiness, or professional conduct (Measure lacks clarity). ■ No assignment of responsibility for equipment losses (driver, weather/act of God, theft, equipment poorly maintained by carrier) – (Inadequate scale) ■ No response category to reflect drivers who experienced no losses (biased measure).
915 – Falsified Employment Application	Valid	Invalid	■ Fails to record employees who submitted true employment applications (measure is biased)
926 – Log Violation	Valid	Invalid	■ Fails to record employees who complied with log procedures (measure is biased)
935 – Company Policy Violation	Invalid	Invalid	■ Measure fails to identify category of policies violated (e.g. customer relations, co-worker relations, professional conduct, safety) – (measure lacks clarity) ■ Measure is likely to be defined differently by carriers according to individual company policy sets (Inconsistent scale). ■ Alternative positive ratings are unavailable (measure is biased)
957, 959, 961 – Unauthorized Equipment Use/ Passenger/ Use of Company Funds	Valid	Invalid	■ Measures contains no indication of culpability (Inadequate scale) ■ Measure fails to identify drivers who routinely complied with equipment, passenger, and company fund polcies (Measure is biased) ■ All 3 measures rely on company standards that likely vary (scales are inconsistent/undefined).

Codes 913 Cargo Loss and 917 Equipment Loss are not clearly related to a category of employment history and thus lack content validity. These measures are inadequate because

they fail to reflect drivers' culpability for cargo and equipment losses. These measures are biased because they focus on a single idiosyncratic event and fail to record drivers who have experienced no equipment or cargo losses.

Code 915 – falsified employment application is clearly related to employee trustworthiness or professional conduct and has content validity. The measure is biased because it fails to record employees who submitted accurate employment applications.

Code 926 – log violation is related to professional conduct. Code 926 is biased because it records only drivers who have violated Federal Motor Carrier Safety Regulations, part 395. It is unclear from the instructions whether this violation relates to the driver's professional conduct, the operation of the truck, or the driver's recordation of that conduct in keeping the log. This problem will likely create inconsistent measurement.

Code 935 – company policy violation lacks content validity because the instructions state that the driver "violate(d) company policies and procedures" and that "this code to be used only when the other selections in this section do not indicate the company policy violated." Because *[redacted]* fails to identify relevant company policies and employment history topics carriers are unlikely to understand when to use this code and even less likely to discern which policies may or may not be reflected by code 935. *[redacted]*s instructions are akin to a teacher telling a student to color a sheet of paper using any color other than the ones other students have used without permitting the student to see other students papers. The best the student can do is guess.

Codes 957 – unauthorized equipment use, 959 – unauthorized passenger, and 961 – unauthorized use of company funds are all reasonably related to professional conduct or employee trustworthiness and have content validity. The measures are inadequate because they fail to reflect culpability. The measures are biased because they fail to identify employees who complied with previous employers' policies regarding equipment use, passengers, and company funds.

Safety: Code 929 – failed to report accident and code 938 – unsatisfactory safety record relate to driver safety. Code 929 lacks direction regarding whether prospective employers should consider the code relevant to safety or professional conduct. Such direction would improve the codes content validity. Code 929 provides no response space for previous employers to indicate the context of the reporting failure and the context of the accident itself. These shortcomings render code 929 inadequate. Code 929 provides no response category

that identifies drivers who have not had accidents. Although potential employers could learn more about a driver's safety record from sections c through m of the termination record form, code 929 itself reports nothing directly about an employee's safety practices. The work record section's clarity could be improved by moving code 929 to sections c through m and by inquiry about accident reporting for each corresponding accident.

Code 929 focuses on a single idiosyncratic event, a one-time failure to report an accident, and is biased. Code 929 – failed to report accident is invalid with respect to measurement.

Table 2.5: Safety

Work Record Measure	Content Validity	Measurement Validity	Explanation
929 – Failed to report accident	Invalid	Invalid	■ No instructions relating measure to safety or professional conduct (measure lacks clarity) ■ No indication of responsibility for failure to report accident (inadequate scale) ■ No response provided for drivers who report accidents according to guidelines (measure is biased) ■ No response provided for drivers who have not had accidents (measure is biased)
938 – Unsatisfactory Safety Record	Valid	Invalid	■ Safety standards likely vary from company to company (Inconsistent scale). ■ Threshold for "Unsatisfactory" is undefined (unclear scale) ■ Alternative positive rating is not available (biased measure) ■ Measure focuses on a single event (biased measure).

Code 938 – Unsatisfactory Safety Record is valid with respect to its related content, driver safety. The directions state, "Driver did not meet company safety standards." The reliance on company-specific standards is likely to introduce varied scales into the measure and decrease its measurement validity. There is no threshold definition for "Unsatisfactory" other than the driver not meeting standards. The implicit threshold is a single event in which the

driver fails to meet a company safety standard regardless of the driver's culpability for that failure. The measure's inadequacy undermines its measurement validity.

There is no positive categorical response available to represent drivers who routinely met company safety standards. This bias renders the measure invalid.

The [redacted] Termination Record form contains two codes, Code 944 – personal contact requested, and Code 999 – other that have no content validity. Code 944 – personal contact requested is not a measure and thus the validity concepts do not apply. However by including code 944 along with measures *[redacted]* has compounded the disorganized contents of its work history section. Because many previous employers use only a single code to reflect an entire employment history this code may effectively deprive potential employers of valuable information.

Table 2.6: General Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
944 – Personal Contact Requested	Not Applicable	Not Applicable	■ This code is not a measure
999 – Other	Invalid	Invalid	■ Allows for measure to relate to any category of performance none of which are identified either previously or by the previous employer recording "other" (measure lacks clarity) ■ Provides no information about driver character or event histories (scale is absent)

For code 999 – other, *[redacted]*s instructions are "anything other than items listed above." These instructions are so broad that the code may completely lack a relationship to employment history. The lack of categorical direction renders the measure invalid with respect to content validity. The code is not biased nor is it inadequate. The general nature of the code provides prospective employers with no indication of whether the "other" employment event or driver characteristic is positive, negative, or neutral.

In the work record section of the [redacted] Termination Record form there are nine measures (931 – Quit Under Dispatch, 933 – Quit/Dismissed During Training or Orientation, and codes 950-956) that identify single idiosyncratic events related to the context of employment termination. None of these codes provides responses reflecting drivers who did not terminate their employment under the circumstances described. The [redacted] Termination Record form

instructs previous employers to circle only one code among codes 950 through 956, but does not indicate whether doing so then limits a carrier's responsibility to use codes 901 through 944 and code 999. This confusion and lack of direction could further erode carriers' abilities to use the codes that apply to specific drivers and thus undermines the measure's scale and its measurement validity.

By way of illustration, I have produced an alternative temination record form work record section that I present as Appendix B. The ordinal categorical responses provide for greater adequacy and reduce measurement bias. If carriers used this form, they could record drivers' experiences and characteristics as they were reflected throughout employment histories. The comprehensive scales provide response categories for unsatisfactory, satisfactory, and exceptional driver characteristics, and carriers could positively identify when they have not observed a characteristic or event such that it cannot be recorded on behalf of a driver.

The twenty measures require fewer dichotomous identification choices than the current *[redacted]* work record form and the superior organization, with contents grouped among three topics, should facilitate accurate reporting. Given the current low use of work record codes, with 94 percent of carriers using only one work record code per termination record form, *[redacted]* would greatly enhance the reliability of its employment histories by creating a better organized employment history report with adequate and unbiased response scales.

Carriers could include summary measures for each topic on the alternative form. Prospective employers would have information from the individual measures within each topic such that they could evaluate whether summary topical codes of exceptional, meets expectations, or does not meet expectations comported with the individual codes recorded. *[redacted]* could maintain its separate section for accident reporting with additions regarding accident reporting by drivers.

In concluding this evaluation of the [redacted] Termination Record form and its instructions all of the deficiencies identified up to this point are based on the form's presentation, design, and instructions. None of the evaluations of content and measurement validity require a statistical analysis of the employment histories produced by the [redacted] Termination Record form. Rather, these design matters are typically attended *ex ante* by researchers and recorders.

Section 3:
Analysis of Carrier *[redacted]* Form Profiles

In this section I move from a critique of the design and presentation of and instructions for the [redacted] Termination Record form to an evaluation of how consistently suppliers of statements to *[redacted]*'s database, previous employers, use the termination record form to reflect employment events and driver characteristics. The forthcoming analysis definitively reveals that *[redacted]*'s suppliers, carriers, use the termination record form codes inconsistently and the analysis strongly reflects that a majority of carriers do not have a shared understanding of when and how to use specific codes. Even a cursory non-statistical analysis identifies hundred of carriers whose code use lays beyond reasonable bounds. For example, 102 carriers have submitted over 80,000 termination record forms without once using code 901 satisfactory. Even with the articulated shortcomings in its design and the validity problems discussed in the previous section *[redacted]*'s managers could have and can readily assess the consistency with which motor carriers understand and employ the termination record form, and can identify any individual motor carrier or group of motor carriers who utilize termination record forms in an unusual manner. *[redacted]* could have and can identify carriers whose termination record forms lie beyond a reasonably expected range of code use by comparing the profiles of code usage from any single motor carrier with the profile of termination record form codes produced by a sample of one thousand or more termination record forms from the population of all *[redacted]*'s termination record forms.

This evaluation does not depend upon any termination record form measure being clear, unbiased, or adequate (i.e. having content and measurement validity). This evaluation assesses carriers' shared understanding of the *[redacted]* codes as that understanding is reflected by their code use on termination record forms they have voluntarily submitted to *[redacted]*. If *[redacted]*'s data suppliers, motor carriers, have a shared understanding of the termination record form codes and consistently use those codes to reflect driver experiences and characteristics then the profiles of the individual data suppliers' (i.e. carrier companies) termination record form code usages should not be grossly different from the code usages in the random sample of termination record forms.

We can assess the carriers' use of termination record forms by developing a range of likely percentages for carriers based upon the code usages observed in a random sample of 5000 termination record forms (hereafter the TRF sample). Statisticians refer to this range of

likely outcomes as a Confidence Interval. A confidence interval uses percentages observed in a population sample to provide a lower boundary and an upper boundary of the percentages that we would expect to observe in future or comparative samples. In this case our comparative samples are the profiles of code use for 3604 carriers that have contributed 30 or more termination record forms to *[redacted]*.

In order to understand confidence intervals a preliminary example may be helpful. If we say that a sample of 1000 school children reports that ten percent of children age 10 are taller than 60 inches tall and that the ten percent value from the sample has a 95 percent confidence interval ranging from eight percent to twelve percent, then we are saying that in 19 out of 20 future or comparative samples of 1000 school children we expect that the proportion of children who are more than 60 inches tall will range within this confidence interval which has a lower bound of 8 percent to an upper bound of 12 percent. Notice we do not expect any single sample to estimate that exactly 10 percent of children are more than sixty inches tall. Rather we expect that for every twenty samples, 95% or 19 of them will report that the percentage of children who are more than 60 inches tall ranges from a lower bound of eight percent to an upper bound of twelve percent. The twentieth sample can then be thought of as an outlier. This means that we have no specific expectation about the percentage of children estimated to be more than sixty inches tall in that 20^{th} sample other than that the proportion is likely to be less than 8 percent or greater then 12 percent.

If a second researcher reports ten subsequent samples of 1000 school children age 10 and four samples, 40% of the samples, estimate that more than 12% of children are taller than 60 inches, then we would estimate that three of the four samples lie beyond the expect range of percentages. If we check to ensure our initial sample is taken properly, then we would inquire about the second researcher's sampling procedure. If he reports that he sampled basketball camps then that qualification could explain why the children in his samples are reported as taller.

Before analyzing the [redacted] Termination Record form and carriers' use of individual codes within the termination record form, I develop what a confidence interval is and how it works. I first outline the components or ingredients in the confidence interval and then I walk through how it is used to evaluate reports from subsequent samples or sources.

The ingredients of a confidence interval include:

1. A population of interest
2. The property of interest from that population
3. A sampling procedure and the sample size
4. Estimated population proportions from the sample
5. Estimated standard errors for each proportion of interest
6. A confidence level

The population of interest in *[redacted]* is the population of termination record forms completed by carriers to represent drivers' employment histories. Notice that the population is not drivers or carriers but completed termination record forms. The TRF sample of 5000 termination record forms provides a representative sample of the population of 3.5 million termination record forms. Three thousand six hundred three carriers contributed 30 to 106,272 termination record forms to this population. On average these 3603 carriers contributed 999 termination record forms and these termination record forms comprise over 97% of the termination record forms in the population. In this analysis, we are interested in how carriers have used termination record forms to produce drivers' employment histories.

The property we are interested in from the population of termination record forms is proportion of forms that use each of the individual work record codes (i.e. codes 901 through 999). In other words what percentages of termination record forms have code 901 satisfactory marked or code 938 unsatisfactory safety record marked? We could be interested in this property for any number of reasons. For this analysis we will want to compare the percentages from the population of termination record forms, that is the profiles of code use on termination record forms, to the proportions of codes marked by individual companies.

The sampling procedure was to randomly sample 5000 termination record forms from all the 3.5 million completed forms available in the population of termination record forms. A random sample does not mean that work record forms were selected in an *ad hoc* or helter skelter fashion. A random sample means that every termination record form in the population

of termination record forms had an equal probability of being selected for the sample. If there were 500,000 termination record forms in the available population and we sampled 5000, then the probability of being sampled was 0.01. If there were 100,000 termination record forms in our sample of 5000 termination record forms then the probability of being randomly sampled was 0.05 or 5 percent. A random sample is the best sampling procedure to ensure that the sample is representative of the population (Judd, Kidder, and Smith, 1991).

Choosing a sample size is important because larger samples will provide greater precision in estimating population proportions. Larger sample sizes generate more precise estimates of population proportions (Babbie, 2004). Most survey samples of the population of the United States have a size of 900 to 1200 respondents or observations and these sample sizes result in confidence intervals that range from 3.2% below a sample proportion of 50% to 3.2% above a sample proportion of 50%. By using a sample size of 5000, we have increased our precision. In contrast to confidence intervals that are 3.2% above and below a sample proportion of 50%, our confidence intervals range from 1.4% above a sample proportion of 50% and 1.4% below a sample proportion of 50%. Notice, the sample of 5000 [redacted] Termination Record forms is four times larger than most samples of the United States population. The precision and "representativeness" of random samples does not depend on the size of the population but rather is a function of the sample size.

The fourth step in developing confidence intervals is to estimate population proportions from the TRF for each work record code. For the *[redacted]* work record codes the sample of 5000 produced an estimated population proportion for each code and these are reported below in table 3.1.

Table 3.1: Estimated Population Proportions from Sample of 5000 Termination Record Forms

Code	Proportion From Sample of 5000 TRF
901 Satisfactory	
912 Excessive Complaints	
924 Late Pick Up/Delivery	
926 Log Violation	
928 No Show	
931 Quit Under Dispatch	
933 Quit/Dismissed During Training/Orientation	
935 Company Policy Violation	
938 Unsatisfactory Safety Record	
944 Personal Contact Requested	
999 Other	

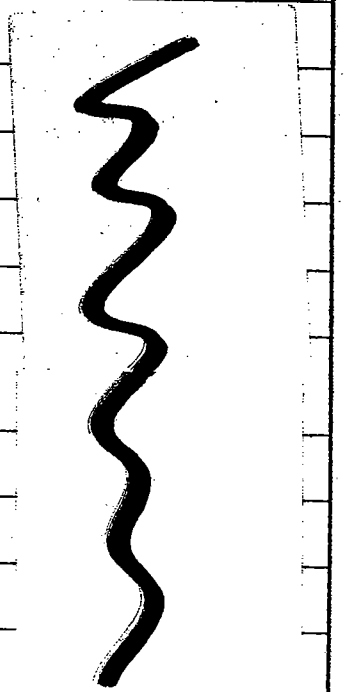


Table 3.1 lists the proportions we observed of code use observed from the random sample of 5000 termination record forms. For example we observed that carriers marked [redacted] drivers as code 901 satisfactory on the termination record form. This [redacted] codes used translates to an estimated population proportion of [redacted], and we interpret this by stating that among the three million termination record forms in the population we estimate that [redacted] of them are marked using code 901 satisfactory. For this analysis, I excluded those codes that were marked on less than 0.5 percent (25) of the termination record forms.

Although a sample of 5000 provides relatively precise estimates of population proportions, we now want to develop the interval in which we expect comparative sample proportions to fall. Remember we don't expect subsequent samples of 5000 termination record forms to duplicate our initial sample, but we do expect them to be within a defined range of percentages.

The next step in our procedure, noted as step 5 on page 27, is to estimate the standard error for each proportion. The standard error for sample proportions is the standard deviation of the proportion adjusted for the sample size.

The standard deviation is the average amount that sample observations differ from the sample mean. To calculate the standard deviation we follow these steps:

1. In our sample carriers marked code 901 satisfactory to denote satisfactory drivers, and these drivers were given a value of one (1) on the termination record forms in our sample. Drivers not marked as one were coded as zero (0) on code 901 on the termination record forms in our sample. The TRF sample percentage was [redacted] meaning that [redacted] termination record forms were coded as 1 on code 901 satisfactory, and [redacted] of termination record forms in the sample were coded as 0 on code 901 satisfactory.

2. The mean score of code 901 is equal to the percentage of termination record forms marked as code 901, which is the estimated population percentage of [redacted]

3. To estimate the standard deviation from the sample percentage, we multiply the TRF sample proportion, [redacted] by its counterpart, 1 – (sample percentage/100) (Diamond and Jeffries, 2001). These two terms represent the percent of drivers coded satisfactory and the percent not coded as satisfactory on our 5000 termination record forms. Then we take the square root of this product to give us the standard deviation or average distance from the estimated population percentage.

4. The formula for the standard deviation of a proportion is:

   Standard deviation = SQRT [Percentage x {1 – (Percentage/100)}]

Now we want to move from the standard deviation of a percentage to the standard error of a sample proportion. To do this requires that we adjust the standard deviation we just estimated to reflect the sample size. Remember, larger sample sizes provide greater precision, which means that larger sample sizes reduce the size of the standard deviation. To make this adjustment, we divide the standard deviation by the square root of the size of our sample. Our

sample is 5000 termination record forms so we will divide the standard deviation of the sample proportion by the square root of 5000. Our formula now is:

Standard error of a proportion = SQRT {Percentage x (1 – Percentage)}/ Sample Size}

With the standard error we can construct a confidence interval that is centered on the TRF sample proportion, which is the estimated population percentage. To continue this process we should select a *Confidence Level*. The confidence level provides a level of precision for our expectations about the samples or groups we wish to compare to our random and representative population sample.

A confidence level has a specific interpretation. If we choose a 95% confidence level, we are saying that we expect 19 of 20 (95%) of subsequent samples of 5000 or more observations will fall within our confidence interval. With a 90% confidence level we expect 9 of 10 samples of 5000 or more termination record forms to fall within the range of our 90% confidence interval. With a 99% confidence interval we expect 99 of 100 comparative samples to fall within the range of our confidence level. We are not saying that we are 99% sure of our estimate. We are saying that given our sample proportion, its standard deviation, and our sample size, we expect the corresponding percentage of comparative samples of that size to be within our confidence interval.

The confidence interval permits us to identify outlier samples in comparison to the population percentage estimated by the TRF sample, which is representative of the population of termination record forms. Generically, outlier samples are samples in which the reported values lie far beyond an estimated population proportion or that lie beyond the range of values from most of the samples. Statistically outlier samples can be defined as samples in which one or more reported values lay two or more standard errors outside the confidence interval. In this analysis the TRF sample generates the estimated population proportion and the termination record forms contributed by each of the 3603 carriers to the population of 3.5 million termination record forms are the subsequent comparable samples.

For every confidence level there is a value from a statistical distribution, called the Normal Distribution, that tells us the range of sample percentages that will fall above and below the estimated population percentage for which we can attribute the deviation from the population percentage to to sampling variation. We can infer that sample perentages that fall

within this range are expected from sampling variation and are not caused by systemic distortions.

The range around an estimated population percentage that we must deviate to ensure that the percentage of comparative samples that fall within the confidence interval is equivalent to our confidence level. For every confidence level, we can select a corresponding "Z-score" from which we calculate the confidence interval. We can think of a Z-Score as a "distance score" that tells you how from the TRF sample percentage to estimate the boundaries of comparative samples that should fall within a range estimated by the TRF sample percentage and its standard error. Because sample sizes with 30 or more observations enable us to use the normal distribution we can use the Z-scores, or distance scores, to estimate how far from our sample proportions we must deviate to create any confidence level (Diamond and Jeffries, 2001; Smith, 1985). The equation for creating a confidence interval then is:

Confidence Interval = Sample Percentage +/- {Distance Score x Sample Standard Error}

The confidence interval is centered upon the TRF sample proportion. The confidence interval has an upper bound equal to the TRF sample percentage plus the product of the distance score (Z-score) times the TRF sample percentage's standard error. The Confidence interval has a lower bound equal to the TRF sample percentage minus the product of the distance score (Z-score) times the TRF sample percentage's standard error.

The normal distribution provides consistent parameters for us to develop confidence intervals. When using the normal distribution we expect that 68.25% of all subsequent samples (i.e. those samples drawn from the same population and of the same sample size) will fall within one standard error above or below our sample percentage. Ninety-five percent of comparative samples will fall within 1.96 standard errors of our sample proportion, and 99% of comparative samples will fall within 2.575 standard errors above or below our sample proportion (Diamond and Jeffries, 2001; Babbie, 2005). Because *[redacted]* has not provided a its own statistical analysis of termination record form work record codes I chose the confidence level, and corresponding Z-score, that generated the widest confidence intervals and minimized the number of carriers likely, in the absence of systemic distortions, to be identified as deviants or outliers.

For example for termination record forms that use the code 901 satisfactory, we have a TRF sample percentage of [redacted] and a standard error of [redacted]. If we create an interval that is the sample proportion plus or minus 1 standard error than our interval is

68.25% Confidence Interval = [redacted] +/- (1) [redacted] = [redacted] to [redacted]

If we deviate from the TRF sample percentage by plus or minus one standard error then we would expect [redacted]% of comparative samples to have sample percentages for code 901 satisfactory within the range of [redacted]% to [redacted]%. To create a 95% Confidence Interval, in which we expect 19 of 20 comparative samples to fall within, we have to use the distance score of [redacted] so that now our equation is:

95% Confidence Interval = [redacted]% +/- ([redacted] = [redacted]% +/- 1.4% = 56.1% to 58.9%

This 95% confidence interval has a lower bound of [redacted] and an upper bound of [redacted]%. It is wider than our [redacted] confidence interval. We interpret this [redacted]% confidence interval by stating that we expect 19 of 20 subsequent samples of 5000 termination record forms will have sample proportions for code 901 satisfactory that lie between [redacted]% and [redacted]%. We expect one of twenty samples of 5000 termination record forms will lie outside the interval [redacted]% to [redacted]%.

Before moving to an evaluation of the carriers' use of [redacted] termination record forms, two final points about confidence intervals are in order. First, higher confidence levels lead to wider confidence intervals. A 68% confidence interval ranges from one standard error below the sample proportion to one standard error above the sample percentage. A 95% confidence interval ranges from 1.96 standard errors below the sample percentage to 1.96 standard errors above the sample percentage. A 99% confidence interval ranges from 2.575 standard errors below the sample percentage to 2.575 standards error above the sample percentage.

Second, larger sample sizes create smaller standard errors of the sample percentages and lead to smaller or tighter confidence intervals. Small sample sizes lead to larger or wider confidence intervals. A sample percentage of 40% from a sample of 1000 termination record forms has a standard error of 1.55%. A sample proportion 0f 40% from a sample of 30 termination record forms has a standard error of 8.94%. When estimating confidence intervals we can ensure comparability by using the smallest sample size from the comparative samples to create the confidence intervals. Larger samples will fall within the confidence intervals created

using smaller samples, but smaller samples may or may not fall within confidence intervals using larger samples.

I have used the random representative sample of 5000 termination record forms to generate the estimated population percentages. The sample of 5000 termination record forms produced the estimated population percentages for all [redacted] termination record forms used to produce drivers' employment histories. The records of code use by 3603 carriers are then compared to these estimated population percentages. The 3603 carriers each has a profile that reports 1) the total number of termination record forms submitted by the carrier and 2) the frequency (i.e. the number of termination record forms) with which a carrier used each individual work record code. To obtain individual carrier code use percentages that could be compared to the TRF sample, I divided the frequency of code use by the total number of termination record forms submitted by that carrier, and then multiplied the proportion by one hundred. Each carrier's code use is then a sample that can be evaluated against the TRF sample of 5000 termination record forms and the confidence intervals generated around its percentages of code use.

Carriers submitted a minimum of 30[redacted] termination record forms and a maximum of [redacted] termination record forms. The mean number of termination record forms submitted by a carrier was [redacted]. The total number of termination record forms profiles in the sample of 3603 carriers is over 3.5 million.

Confidence Intervals & Evaluation of Carriers' Code Use. To evaluate how carriers use the work record codes 901 through 999 in the termination record form, I used the estimated population percentages that are presented again here as they were in table 3.1. To estimate the standard errors for these percentages, I used 30 as the sample size for all the codes. This ensures that all the confidence intervals are comparable for carriers whose profile included the minimum of 30 termination record forms. This assumption biases the evaluation in [redacted] favor and means that many carriers who have used codes on an unusually small or larger percentage of termination record forms will go undetected because of the generous confidence intervals. Overall, we expect 99 of 100 the carriers in the sample of 3603 carrier profiles to fall within the confidence intervals.

Table 3.2: Termination Record Form Work Record Codes and Estimated Population Percentages

Code	Proportion From Sample of 5000 TRF	Standard Error of the Percentage (Sample size = 30 trf's or more)
901 Satisfactory		
912 Excessive Complaints		
924 Late Pick Up/Delivery		
926 Log Violation		
928 No Show		
931 Quit Under Dispatch		
933 Quit/Dismissed During Training/Orientation		
935 Company Policy Violation		
938 Unsatisfactory Safety Record		
944 Personal Contact Requested		
999 Other		

With these estimated population percentages and their standard errors for samples of size 30 or greater, I generated 99% confidence intervals for each code. These confidence intervals have Z-scores, or distance scores, of 2.575, which is from the Normal Distribution. We interpret 99% confidence intervals to mean that we expect 99 of 100 comparable samples will fall within the interval created by adding and subtracting the product of 2.575 times the standard error of the proportion to and from the sample proportion. I have converted proportions back to percentages and this conversion does not affect the results or evaluation.

After generating each confidence interval I identified how many carriers exceeded the lower or upper boundary of the confidence interval. For code 901 satisfactory I identified how many carriers' use of each code were below the lower boundary. For the remaining codes I identified how many carriers' use of each code were above the upper boundary. Among 3603 carriers, we expect [redacted] carriers (99% of carriers) to lie within each confidence interval for

each work record code. We expect [redacted] to lie either above or below the confidence interval with approximately half, [redacted] carriers, exceeding the interval, and approximately half, [redacted] carriers, lying below the interval. In cases where the interval has a lower bound below zero, I have conservatively assumed all one percent of carriers, [redacted], will exceed the confidence interval because they cannot fall below it.

I will first deal with code 901 satisfactory. Table 3.3 presents the estimated population percentage for code 901 satisfactory, the confidence interval for that percentage, the number of carriers whose use we expect in the absence of systemic distortions to fall below the lower bound of the confidence interval, and the number of carriers observed below the lower bound of the confidence interval.

Table 3.3: Confidence Interval for Code 901 Satisfactory

Code	Percentage From Sample of 5000 TRF	99% Confidence Interval	Number of Carriers expected to exceed Lower Bound of confidence interval	Number of Carriers Observed Exceeding Lower Bound of Confidence interval
901 Satisfactory*	[redacted]	[redacted]	[redacted]	[redacted]

As table 3.3 reflects [redacted] of termination record forms in the sample of 5000 termination record forms were coded as satisfactory. Subtracting and adding the product of Z-score of 2.575 times the standard error, [redacted], from the estimated population percentage, [redacted]% generated the confidence interval for this code. The lower bound of this confidence interval is [redacted]. The upper bound of this confidence interval is [redacted]. We expect [redacted] carriers to have identified less than [redacted] of the drivers coded on their termination record forms as satisfactory by using code 901. In the 3603 carrier samples we observe that [redacted] carriers use code 901 satisfactory on then [redacted]% of their termination record forms.

To evaluate this result recall that we expect one-half of one percent of carriers to have profiles falling below the confidence interval. We observe [redacted] of 3603, or [redacted], carriers using code 901 satisfactory at rates that fall below the lower bound of the confidence interval. In other words *there are 33 times as many carriers below the confidence interval's lower bound as we expected* given our estimated population percentage, our 99% confidence level (which

widens the confidence interval to bias the interpretation in favor) and our assumption that every carrier contributed only 30 termination record forms (which further widens the confidence interval to bias the analysis in favor). From the result that carriers lay below the confidence interval we can infer that % of carriers contributing termination record forms to in ways that are systematically different from ways the estimated population percentages for termination record forms. Although a specific cause cannot be isolated, there are systemic distortions in carriers' use of code 901 satisfactory.

These statistically significant deviations, occurring among one-sixth of carriers in the carrier sample, are based upon comparing individual carriers' use of termination record forms with the TRF sample. The producers of both samples are carriers who have been drivers' previous employers. We are not comparing carriers' use of termination record forms to drivers or to driver's self-reports about employment history. In other words, this analysis compares apples to apples or carrier produced termination record forms to carrier-produced termination record forms.

The frequency with which carriers deviated from the estimated population percentage, with one in six carriers falling below the lower bound of the confidence interval should have been plainly evident to In the sample of carriers carriers) used code 901 satisfactory zero (0%) of the time! Although these carriers represent slightly less than three percent of all carriers they represent nearly *six times* the expected frequency of carriers using code 901 satisfactory on zero to thirty-four percent of their termination record forms. These carriers submitted between and termination record forms. Collectively these carriers submitted termination record forms never using code 901 satisfactory. Any casual observer, including managers, could identify these carriers without performing a statistical analysis.

The decision to use a 99% confidence interval and to assume only 30 termination record forms per carrier for purposes of estimating the confidence intervals means that we are very likely underestimating the number of carriers who fall below the confidence interval. For subsequent evaluations of individual codes, I will be underestimating carriers' code usages that fall above the upper bound of the 99% confidence intervals.

Table 3.4 presents the estimated population percentage for work record codes 912 through 999, the confidence interval for each percentage, the number of carriers we expect to

exceed the upper bound of the confidence interval and the number of carriers observed to exceed the upper bound of the confidence interval.

Table 3.4: 99% Confidence Interval for [redacted] Work Record Codes 900 through 999

Code	Percentage From Sample of 5000 TRF	99% Confidence Interval	Number of Carriers expected to exceed Upper Bound of confidence interval	Number of Carriers Observed Exceeding Upper Bound of Confidence interval
912 Excessive Complaints	[redacted]	[redacted]	[redacted]	[redacted]
924 Late Pick Up/Delivery	[redacted]	[redacted]	[redacted]	[redacted]
926 Log Violation	[redacted]	[redacted]	[redacted]	[redacted]
928 No Show	[redacted]	[redacted]	[redacted]	[redacted]
931 Quit Under Dispatch	[redacted]	[redacted]	[redacted]	[redacted]
933 Quit/Dismissed During Training/Orientation	[redacted]	[redacted]	[redacted]	[redacted]
935 Company Policy Violation	[redacted]	[redacted]	[redacted]	[redacted]
938 Unsatisfactory Safety Record	[redacted]	[redacted]	[redacted]	[redacted]
944 Personal Contact Requested	[redacted]	[redacted]	[redacted]	[redacted]
999 Other	[redacted]	[redacted]	[redacted]	[redacted]

For the remaining work record codes we expect no more than [redacted] carriers to have profiles in which they use an individual code on a greater percentage of termination record forms than has been indicated by the upper boundary of the 99% confidence interval. We observe the following:

1. For code 912 Excessive Complaints, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]%. This is **8.7** times greater than expected.

2. For code 924 Late Pick Up/Delivery, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **16.2** times greater than expected.
3. For code 926 Log Violation, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]%. This is **14.2** times greater than expected.
4. For code 928 No Show, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]%. This is **9.4** times greater than expected.
5. For code 931 Quit Under Dispatch, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]% . This is **12.9** times greater than expected.
6. For code 933 Quit/Dismissed During Training, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **5.6** times greater than expected.
7. For code 935 Company Policy Violation, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **8.6** times greater than expected.
8. For code 938 Unsatisfactory Safety Record, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **4.0** times greater than expected.
9. For code 944 Personal Contact Requested, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **16.1** times greater than expected.
10. For code 999 other, [redacted] carriers are above the upper boundary of the confidence interval of [redacted]. This is **6.4** times greater than expected.

For all ten of these work record codes and code 901 satisfactory the carrier profiles from the sample of 3603 carriers unmistakably reveal that substantial numbers of carriers systematically use work record codes in ways that dramatically deviate from reasonable statistical expectations. This analysis does not provide the means to impute a cause for these statistically significant deviations and systemic distortions from the expected range, but I cannot stress enough that these deviations are very substantial and very frequent. This analysis very strongly reflect that [redacted] suppliers, drivers' previous employers, share little if any understanding of the [redacted] termination record form codes and under what circumstances to use which codes.

Even after making every effort to estimate the widest possible confidence intervals we observe a minimum of [redacted] of [redacted] carriers using codes beyond the expectation. Not a single code of the eleven codes evaluated has carrier profile percentages that fall within the confidence intervals estimated with the sample of 5000 termination record forms.

One response to this analysis could be that a minority of carriers, perhaps 15% of carriers, could creating these frequent and substantial deviations by always using codes based on individual deviant practices.[2] This is not the situation. Using the SPSS software I identified carriers that fell outside of the confidence interval for one or more of the codes in tables 3.3 and 3.4. Among [redacted] carriers [redacted] fell outside of one or more confidence intervals leaving only [redacted] that were consistently within the confidence intervals. In other words [redacted] of carriers in the carrier sample demonstrated deviant coding for at least one work record code. This assessment understates carriers' coding problems because it is based on the widest possible confidence intervals. A carrier-by-carrier analysis would reveal that an even larger percentage of carriers fall beyond estimated confidence intervals because such an analysis would adjust for the number of termination record forms submitted by each carrier in developing the confidence intervals.

If I were a business manager identifying sources of information that lay beyond statistically reasonable expectations I would recommend using a tighter 95% confidence interval that would provide an expectation of 19 of 20 carriers coding within the confidence interval and I would adjust the confidence intervals for the number of termination record forms submitted by individual carriers rather than using a minimum of 30 which results in generous confidence intervals. These two adjustments would reveal that more than [redacted] carriers used one or more work record codes with frequencies that exceeded the bounds of estimated confidence intervals.

This analysis reflects the lack of content and measurement validity of carriers' use of the termination record form and employment histories. With respect to content validity carriers' wide ranging use of codes reflects a lack of shared understanding that goes beyond what we would expect from the variable experiences drivers are likely to have. With respect to measurement validity the most problematic information comes from the large number of carriers who infrequently code drivers as "satisfactory" and the large number of carriers who

[2] "Deviant" here refers to statistical departures from the expected range and is not meant to provide a normative connotation.

frequently use codes such as 926 log violation, 933 quit/dismissed during training or orientation, and 938 unsatisfactory safety record. If these measures were valid and representative of drivers' employment histories one would have to inquire how companies could remain viable with less than one-third of their employees being satisfactory and in [redacted] observations carriers have operated with no satisfactory drivers.

Section 4:

Reliability in the *[redacted]* Form Work Record Section

The [redacted] Termination Record form's "Work Record" section has twenty-seven items that together present employment histories of drivers employed by carriers who contribute termination record forms to *[redacted]*. An important consideration for *[redacted]* and potential employers is whether the employment histories presented are ***reliable*** representations of drivers' employment histories. Reliability is the extent to which a survey, questionnaire, test, or other measurement instrument *consistently* depicts the same events and characteristics (Carmines and Zeller, 1979: Lewis, 1999; Litwin, 1995; Thompson, 2003). A lay definition of reliability is the degree to which repeated measurements, or measurements taken under identical circumstances, will yield the same results. A statistical definition of reliability is that reliability is a measure of the randomness of the measurement process itself. Reliable measures minimize the random or variable component of the measurement process (Lewis, 1999).

An example may bring the concept of reliability into stark relief. Suppose I weigh myself on my bathroom scale one morning and the scale reads 164 pounds. Ten minutes later I step on the scale again and it reads 165 pounds. Is this scale unreliable? Probably not. The one-pound difference may be attributable to a number of factors – perhaps I put my shoes on between the measures, perhaps I ate a large breakfast, or perhaps the reading changed because of how I centered myself on the scale.

Next assume I weigh myself on a second scale the next day and it reads 163 pounds. Ten minutes later I check again and the scale reads 191 pounds! To make sure I didn't rig the scale or misread it, I step on the scale a third time. Now the scale reads 147 pounds! Three readings with a range of 44 pounds over ten minutes cannot be attributed to any of the factors that may have caused the one pound difference the previous day on the first scale. I would infer that a scale that provides three readings with a range of 44 pounds is unreliable. Because this bathroom scale is unreliable it produces inaccurate measures of my weight. Unreliable surveys and questionnaires produce inaccurate information and unreliable employee evaluation forms produce inaccurate employment histories (Brown, 2005; Litwin, 1995).

Before moving to an analysis of the [redacted] Termination Record form's reliability I note the frequency with which carriers fail to use work record codes at all. Before we can rely

[redacted] termination record form. Termination record forms that use none of the work record codes are incomplete and inaccurate. Carriers who do not use the work record codes mean [redacted] cannot produce work histories. Among 5000 termination record forms in a sample of termination record forms submitted after July 1, 1999, carriers used no work record codes (number 901 through 999) for [redacted] termination record forms. In other words [redacted] of termination record forms failed to produce work histories. Even if all twenty-six work record codes number 901 through 961 are inapplicable to a driver's employment history, the carrier has the opportunity to so indicate by using the code 999 – other in the work record section. This initial error rate could have been and can be readily observed by [redacted]

[redacted] instructs carriers to use all codes that apply to drivers' employment histories. Nevertheless the vast majority of carriers use only one code per driver. This has the effect of decreasing the variance of the individual codes and substantially decreasing the overall summed "score" of experiences and characteristics for each driver. Despite [redacted]s claims that it views the work record codes as "issue identifiers" its instructions to check all codes that apply are instructions to produce complete employment histories. In the sample of 5000 termination record forms, only [redacted] termination record forms, [redacted], had two or more 900 codes checked. Among drivers not coded as satisfactory, superior, or outstanding, only [redacted] out of [redacted] had two or more work record codes identified. Among these same [redacted] drivers [redacted] drivers had no 900 codes used and no employment history produced on their behalf by the work record section of the termination record form. Carriers are more likely to represent that drivers had no work experiences, including "other," than they are to identify drivers as having one or more work experiences or driver characteristics. The inference from this data summary is that most termination record forms produce incomplete, and therefore inaccurate, employment histories.

In addition to the [redacted] termination record forms with no 900 codes used, there were [redacted] termination record forms on which employers used only code 944, personal contact requested, or codes 950 through 956, which identify the location and circumstance of termination. These eight codes, taken alone or together, do not provide information that can then be evaluated as an employment history. Code 944 is not a measure but a request. Codes 950 through 956 deal with a final idiosyncratic event during employment but are not sufficiently informative to denote an employment history. Or [redacted] of 5000 termination record form carriers used too few codes for potential employers to denote an employment history.

For a more complete analysis of the termination record form's reliability, I will use a procedure developed by Frederic Kuder and M. Richardson in 1937 (Kuder and Richardson, 1937). This procedure estimates reliability for surveys and forms in which measures are dichotomously identified. Dichotomous measures include categorical responses such as yes/no, agree/disagree, and applies/does not apply. In the sample of 5000 termination record forms provided by *[redacted]*, identified codes were denoted as one (1) and those left blank or not used were denoted as zero (0). The use of the Kuder-Richardson 20 procedure and formula is appropriate for estimating the [redacted] Termination Record form's reliability as it relates to drivers' employment histories (Litwin, 1995; Thompson, 2003). The Kuder-Richardson 20 procedure estimates the consistency with which carriers use the work record section codes to produce employment histories.

The Kuder-Richardson 20 procedure includes a formula, commonly called the KR-20, that estimates data reliability. The product of the KR-20 procedure, which is called the KR-20 coefficient or KR-20 estimate of reliability, ranges from zero to one as an indicator of reliability.[3] KR-20 estimates equal to zero indicate completely unreliable data and KR-20 estimates equal to 1 indicate completely reliable data (Brown, 2005; University of Connecticut, 2005). An example demonstrates how the KR-20 formula is used to calculate the KR-20 statistic.

We begin with ten employees and the history of their employment training. There are six experiences that the employer records. Each employee is scored as a one if he or she has had the experience and zero otherwise. Thus we might have a table of employees and their experience codes that looks like this:

[3] The Kuder-Richardson 20 formula and its resulting reliability coefficient are very closely related to the reliability formula developed by Cronbach and its resulting reliability coefficient, which is commonly referred to as Cronbach's Alpha (Thompson, 2003). KR-20 is actually actually a special case of Cronbach's Alpha for data that are dichotomously coded (Carmines and Zeller, 1979; Thompson, 2003).

Table 4.1: Employee Experience Codes

Employee	1	2	3	4	5	6
Alice	0	0	0	0	0	0
Bob	0	0	0	0	1	0
Carol	1	0	1	1	1	0
David	1	1	1	1	1	1
Edith	1	1	1	1	1	1
Fred	0	0	1	0	0	0
Gail	0	0	1	1	1	0
Henry	0	0	0	1	0	0
Irene	1	0	1	1	1	0
John	0	1	0	1	0	1

Next we sum the codes for each employee going across the row such that there scores can range from 0 (no experiences recorded) to 6 (all experiences recorded). We now have the following scores for each employee:

Table 4.2: Employee Experience Codes & Summed Scores

Employee	1	2	3	4	5	6	Total
Alice	0	0	0	0	0	0	0
Bob	0	0	0	0	1	0	1
Carol	1	0	1	1	1	0	4
David	1	1	1	1	1	1	6
Edith	1	1	1	1	1	1	6
Fred	0	0	1	0	0	0	1
Gail	0	0	1	1	1	0	3
Henry	0	0	0	1	0	0	1
Irene	1	0	1	1	1	0	4
John	0	1	0	1	0	1	3

The employees' scores range from Alice's 0 to David and Edith's 6's. The mean employee score is the sum of the scores, 29, divided by the number of employees, 10. The mean experience score is 2.9. We now subtract the mean score from each employee's individual score. The resulting differences are each employee's deviation from the mean.

Table 4.3: Employee Scores & Deviation from the Mean

Employee	Total	Mean	Deviation
Alice	0	2.9	-2.9
Bob	1	2.9	-1.9
Carol	4	2.9	1.1
David	6	2.9	3.1
Edith	6	2.9	3.1
Fred	1	2.9	-1.9
Gail	3	2.9	0.1
Henry	1	2.9	-1.9
Irene	4	2.9	-1.1
John	3	2.9	0.1

We now want to square each deviation such that the negatives become positive. If we leave the deviations as they are they will simply cancel one another out.

Table 4.4: Employee Scores, Deviation & Squared Deviation

Employee	Total	Mean	Deviation	Squared deviation
Alice	0	2.9	-2.9	8.41
Bob	1	2.9	-1.9	3.61
Carol	4	2.9	1.1	1.21
David	6	2.9	3.1	9.61
Edith	6	2.9	3.1	9.61
Fred	1	2.9	-1.9	3.61
Gail	3	2.9	0.1	0.01
Henry	1	2.9	-1.9	3.61
Irene	4	2.9	-1.1	1.21
John	3	2.9	0.1	0.01

We now sum the column of squared deviations, which totals 40.9. This total is called the "Sum of Squares." We divide this number by the number of employees minus 1 (10-1= 9). The resulting quotient of the Sum of Squares divided by the number of employees less one is 4.544. This quotient is called the variance of the experience scores.

Next we calculate the proportion of employees who the employer recorded as having each individual experience. This proportion is called p. We next will calculate the proportion of employees not having an experience. This is simply 1 – p but can be verified by counting the number of zeroes in each column and dividing this count by 10, the number of employees.

Table 4.5: Employee Experience Codes, Scores & Proportion Having/Not Having an Experience

Employee	1	2	3	4	5	6	Total
Alice	0	0	0	0	0	0	0
Bob	0	0	0	0	1	0	1
Carol	1	0	1	1	1	0	4
David	1	1	1	1	1	1	6
Edith	1	1	1	1	1	1	6
Fred	0	0	1	0	0	0	1
Gail	0	0	1	1	1	0	3
Henry	0	0	0	1	0	0	1
Irene	1	0	1	1	1	0	4
John	0	1	0	1	0	1	3
Proportion having experience	**0.4**	**0.3**	**0.6**	**0.7**	**0.6**	**0.3**	
Proportion not having experience	**0.6**	**0.7**	**0.4**	**0.3**	**0.4**	**0.7**	

We next calculate the product of the proportion of employees having an experience recorded with the proportion not having the experience recorded. These products are:

Table 4.6: Products of Proportions

Task	Proportion Having Experience	Proportion Not Having Experience	Product
1	0.4	0.6	0.24
2	0.3	0.7	0.21
3	0.6	0.4	0.24
4	0.7	0.3	0.21
5	0.6	0.4	0.24
6	0.3	0.7	0.21

We can now move to the KR-20 formula which is as follows:

KR-20 = [#of Codes/ (#of Codes –1)] x [1 – (Sum of "proportion products"/Variance)]

For our example the number of codes is the number of experiences, or 6. The "Proportion Products" are above and sum to 1.35 (0.24+0.21+0.24+0.21+0.24+0.21). The variance is from table 4.4 and is 4.544. Plugging these numbers into the KR-20 formula yields the following:

$$KR\text{-}20 = (6/5) \times [\,1 - (\,1.35/4.544)] = 1.2 \times [1 - 0.297] = 1.2 \times 0.703 = 0.84$$

Our KR-20 estimate of the reliability of the experiences coded by the employer is 0.84. Now we may ask whether this is good or bad. KR-20 ranges from 0 to 1 and increasingly positive estimates reflect increasing reliability (Brown, 2005). Generally, social scientists consider surveys and records with reliability scores above 0.7 to be fairly reliable and those above 0.8 to be reliable (Braumoeller and Goertz, 2000; Brown, 2005; Carmines and Zeller, 1979; Litwin, 1995). A second way of evaluating the KR-20 estimate is to estimate an *error rate* (Braumoeller and Goertz, 2000). This is done straightforwardly and is defined as:

$$1 - SQRT\,(KR\text{-}20) = 1 - SQRT\,(0.84) = 1 - 0.916 = 0.084$$

With the records of employee experiences we have an estimated reliability of KR-20 which equals 0.084 and we have an estimated error rate of 8.4%. We would assess this coding as reliable.

Turning now to the [redacted] Termination Record form we can follow the same procedure to estimate its reliability. Now our sample provides 5000 termination record forms submitted by drivers' previous employers. The characteristics and experiences that are included in the KR-20 estimate include the following:

Table 4.7: Termination Record Form Codes & Frequency of Code Use (from sample of 5000)

Code	Frequency of Use	Percent of 5000 TRF's
912- Excessive Complaints	[redacted]	[redacted]
913 – Cargo Loss	[redacted]	[redacted]
915 – Falsified Employment Application	[redacted]	[redacted]
917 – Equipment Loss	[redacted]	[redacted]
924 – Late Pick Up/Delivery	[redacted]	[redacted]
926 – Log Violation	[redacted]	[redacted]
928 – No Show	[redacted]	[redacted]
929 – Failed to Report Accident	[redacted]	[redacted]
931 – Quit Under Dispatch	[redacted]	[redacted]
933 – Quit/Dismissed During Training/Orientation	[redacted]	[redacted]
935 – Company Policy Violation	[redacted]	[redacted]
938 – Unsatisfactory Safety Record	[redacted]	[redacted]
957 – Unauthorized Equipment Use	[redacted]	[redacted]
959 – Unauthorized Passenger	[redacted]	[redacted]
961 – Unauthorized Use of Company Funds	[redacted]	[redacted]
999 – Other	[redacted]	[redacted]

In reviewing the termination record form's work record section, one may note that other codes are included on the form but not in my analysis. This is for two reasons. First including the codes 901 – satisfactory, 902 – superior, and 903 – outstanding, produces negative KR-20 coefficients.[4] Negative coefficients were also estimated when codes 950 through 956 and code

[4] The interitem correlations among codes 901 through 903 and codes 911 through 999 were negative. Among 51 inter-item correlations, 50 were negative. The only exception was the correlation among code 903 – Outstanding and code 933 – Quit/Dismissed during Training or Orientation. In estimating either KR20 or Cronbach Alpha coefficients one expect positive inter-item correlations. These negative inter-item correlations reinforce the decision to use KR-20 as it requires the least transformation of [redacted]s data.

944 were included. KR-20 ranges from zero to one as an indicator of reliability, and negative products of the formula indicate substantial, systematic problems with the data but cannot be considered estimates of reliability.[5] Therefore codes 901, 902, 903, code 944, and codes 950 through 957 are excluded from the analysis.

A second reason to exclude codes 950 through 956 and code 944 is that these codes do not represent the presence of employee's character or their experience but rather provide limited information about drivers' termination circumstances.

The work record descriptors included in the analysis include the 17 listed in table 7. For every termination record form I added the number of work record descriptors used and this then became the work experience score for each termination record form. This score ranges from zero to seventeen. Despite having seventeen work record codes available, no carrier used more than eight scores on an individual termination record form. Table 8 indicates the frequency of termination record form work record usage.

Table 4.8: TRF Work Record Code Scores

Score	Frequency	Percent
0	[redacted]	[redacted]
1	[redacted]	[redacted]
2	[redacted]	[redacted]
3	[redacted]	[redacted]
4	[redacted]	[redacted]
5	[redacted]	[redacted]
6	[redacted]	[redacted]
7	[redacted]	[redacted]
8	[redacted]	[redacted]

Using the values from the above tables, we can now estimate the KR-20 reliability coefficient. The KR-20 formula is:

KR-20 = [#of Codes/ (#of Codes –1)] x [1 – (Sum of "proportion products"/Variance)]

The values from the tables can be inserted such that the formula becomes:

KR-20 = ([redacted] x [redacted] – ([redacted])] = [redacted] x [1 – [redacted]] = [redacted]

[5] See Thompson (2003) pp. 12-14.

The KR-20 estimate for the reliability of the work record descriptors in the [redacted] termination record form is 0[redacted] well below accepted levels of 0.7 or 0.8 (Brown, 2005; Carmines and Zeller, 1979; Litwin, 1995). The estimated error rate that results from this KR-20 formula is 1 – SQRT ([redacted] which is [redacted] or 60%! Thus the data representing employment histories on [redacted] termination record forms submitted by carriers is very, very unreliable.[6]

The basic analysis of carriers' use of work record codes and the statistical analysis using Kuder-Richardson 20 reveal that the [redacted] termination record form work record section produces very unreliable data and very unreliable employment histories. Because the data are very unreliable they cannot be considered valid in their current form. Data reliability is a necessary but not sufficient condition for validity. This finding does not detract from the analysis detailing the content invalidity and measurement invalidity of the [redacted] termination record form work record section. This finding reinforces the points made about invalidity in the previous sections.

An alternative interpretation of the KR-20 coefficient of [redacted] focuses on statistical "unreliability." Unreliability coefficients are 1 minus any the reliability coefficient, in this analysis KR-20 (Brown, 2005; Lewis, 1999). In this case we would interpret [redacted] as indicating that the data produced by the [redacted] termination record form are 85 percent unreliable. The more conservative interpretation is to use the error rate of 60 percent (Braumoeller and Goertz, 2000).

[6] In a separate analysis, I transposed codes 901 to 903 and codes 911 through 999 into a "satisfaction index." I transposed the codes relating to rehire eligibility into a "rehire index." The Cronbach's Alpha for the two indexes was 0.541. However when the rehire index is compared to the individual identifiers, which [redacted] claims are comprehensive, Cronbach's Alpha falls to 0.28. One contributing factor to this decrease in the estimated reliability is that 94% of carriers use only one or zero work record codes. If termination record forms are incomplete then correlative relationships among the work record codes are suppressed which decreases the data reliability regardless of whether it is evaluated using KR-20 or Cronbach's Alpha.

Section 5: Conclusion

The preceding four sections establish that the [redacted] Termination Record form work record section lacks content validity and measurement validity and that the data produced by *[redacted]* represented as employment histories are unreliable. The statistical analyses presented in sections 3 and 4 are standard and readily available using either SPSS or Stata software and could be conducted using a Microsoft Excel spreadsheet. Even in the absence of these straightforward statistical analyses, *[redacted]* managers could have and should have readily observed anomalies that have been evident in their data since 1999. In this conclusion, I present specific findings from sections 2, 3, and 4.

Section 2 established that the [redacted] Termination Record form lacks content validity and its individual measures are invalid because their scales are inadequate and their response categories are biased. Among twenty work record descriptors, 10 are unclear and lack content validity. Eleven of twenty measures have inconsistent thresholds for identification, seventeen lack adequate response categories, and seventeen are biased. *[redacted]* can easily repair much of this deficiency without adding any additional work for carriers by providing unbiased and adequate response categories and clear directions identifying thresholds for measurement categories.

The lack of clarity in *[redacted]*s work "identifiers" creates confusion about the topics to which the measures relate. *[redacted]* could group identifiers, or measures, according to three to six clearly identified topics and this identification would likely facilitate easier use of the termination record form along with more consistent use of the individual codes. This confusion, which is compounded by the disorganized ad-hoc code presentation in the work record section, renders the work record section invalid with respect to its contents.

The extremely high prevalence of biased and inadequate measures, with 17 of 20 measures being biased and inadequate, renders the individual measures invalid. Because so many identifiers are biased and inadequate and because *[redacted]* instructs carriers to use all codes that apply there can be no doubt that the employment histories *[redacted]* produces from termination record forms are biased and inadequate. These employment histories are inaccurate and cannot be compared to one another. There is no means of assessing the relative frequency of events or the relative extent of driver characteristics.

The analysis in Section 3 reflects the lack of shared understanding among carriers about how and when to use work record codes. Unclear topical relationships among codes (i.e. content invalidity), biased and inadequate individual measures (i.e. invalid measurements), and incomplete and miscoded work record sections (i.e. unreliable data) all contribute to this lack of understanding. Carriers' misuse of work record codes can and should be easily observed and inferred. With over 100 carriers never using code 901 satisfactory and with hundreds of carriers using codes with frequencies that exceed expectations based on sampling and reflect systemic distortions, [redacted]s managers could have readily observed anomalies in its employment histories and they could have addressed these anomalies by changing the termination record form's measures and refining its directions.

The analysis in Section 4 reflects the lack of reliability in the employment histories [redacted] produces using the termination record form work record section. The KR-20 coefficient for the work record descriptors is a paltry [redacted], well below any reasonable expectation above 0.7 or even 0.5. The estimated error rate of 60% is high and suggests that the work record measures are grossly unreliable. This finding is reinforced by the extraordinary frequency with which carriers use only a single work record code including codes relating to load abandonment. Because [redacted]percent of termination record forms have only one or zero work record codes used we can infer that only [redacted]percent of drivers' employment histories contain multiple recordations of driver characteristics and experiences.

As a result of the lack of content validity, the invalid measurements, and the unreliable data produced by termination record forms drivers' employment histories are inaccurate and fail to reflect their range of experiences, the extent of characteristics, and their overall employment experience. Given the multiple and compounding flaws in the [redacted]termination record form work record section, I would assess that employment histories are much more likely to be incomplete and therefore inaccurate than they are likely to be complete and accurate. The [redacted] termination record form work record section is very likely to produce inaccurate employment histories and I can envision no circumstance under which it would provide the maximum possible accuracy.

Finally the termination record forms evaluated in this report were submitted beginning in 1999 through the summer of 2004. The consequences of the termination record form's validity and reliability shortcomings have been evident for five years or longer. The patterns of code misuse and inaccurate or incomplete identification are ubiquitous across measures and persist

over time. Even modest changes in the measures scales and presentation format would yield substantial improvements in accuracy, validity, and reliability.

References

Babbie, Earl. 2005. *The Basics of Social Research, 3rd Edition.* Belmont, CA: Thomson-Wadworth.

Babbie, Earl. 2004. *The Practice of Social Research, 10th Edition.* Belmont, CA: Thomson-Wadworth.

Braumoeller, Bear F. and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44(4):844-858.

Brown, James Dean. 2005. *Statistics Corner: Reliability of Surveys.* University of Hawaii.

Diamond, Ian, and Julie Jeffries. 2001. *Beginning Statistics: An Introduction for Social Scientists.* Thousand Oaks, CA: Sage Publications.

Fowler, Floyd J., Jr. 1995. *Improving Survey Questions: Design and Evaluation.* Thousand Oaks, CA: Sage Publications.

Judd, Charles M., Eliot R. Smith, and Louise H. Kidder. 1991. *Research Methods in the Social Sciences, 6th Edition.* Chicago, IL: Holt, Rinehart, and Winston, Inc.

Lewis, Roger J. "Reliability and Validity: Meaning and Measurement." Presented at the 1999 Annual Meeting of the Society for Academic Emergency Medicine. Boston, MA.

Litwin, Mark S. 1995. *How to Measure Survey Reliability and Validity.* Thousand Oaks, CA: Sage Publications.

Kuder, G.F. and M. W. Richardson. 1937. The Theory and Estimation of Test Reliability. *Psychometrika*, 2:151-160.

Smith, Gary. 1985. *Statistical Reasoning.* Boston, MA: Allyn & Bacon, Inc.

Thompson, Bruce. 2003. "Understanding Coefficient Alpha, Really." In *Score Reliability: Contemporary Thinking on Reliability Issues*, Bruce Thompson, ed. Thousand Oaks, CA: Sage Publications.

University of Connecticut. Kuder-Richardson 20 (KR-20) Formula. Web-based presentation available at www.gifted.uconn.edu/siegle/instrument_reliability_and_validity.htm

United States Department of Labor. "Dictionary of Occupational Titles, Definitions for 904.383-101 Tractor Trailer-Truck Driver and 905.663-014 Truck Driver." Washington, DC.

Appendix A
Employee Evaluation Forms for Rockingham County (VA) Public Schools, Tacoma Public Schools, and the *[redacted]* Form

www.rockingham.k12.va.us/RCPS/Forms/Evals.html

Rockingham County Public Schools
Staff Performance Evaluation Instrument
Bus Driver

Employee's Name: _____ **School:** Choose one:

Name of Evaluator: _____ **Date:** _____

I. Organization and Planning

Evaluation of Performance	Outstanding	Satisfactory	Needs Improvement	Does Not Meet
A. Is punctual and regular in attendance	☐	☐	☐	☐
B. Utilizes work time efficiently	☐	☐	☐	☐
C. Performs all assigned duties promptly, accurately and effectively	☐	☐	☐	☐
D. Demonstrates initiative and good judgment	☐	☐	☐	☐
E. Performs other duties as appropriate and/or directed	☐	☐	☐	☐

Comments:

II. Interpersonal Relations and Communication

Evaluation of Performance	Outstanding	Satisfactory	Needs Improvement	Does Not Meet
A. Creates an inviting and professional atmosphere	☐	☐	☐	☐
B. Maintains a cordial and effective relationship in meeting the public	☐	☐	☐	☐
C. Demonstrates courtesy and professionalism in all communications	☐	☐	☐	☐
D. Appropriately and effectively communicates with co-workers, teachers, administration, students, parents, and the general public	☐	☐	☐	☐

Comments:

III. Professional Responsibilities/Qualities

Evaluation of Performance	Outstanding	Satisfactory	Needs Improvement	Does Not Meet
A. Maintains effective working relationships with other employees; works effectively as a team member; maintains positive attitude.	☐	☐	☐	☐
B. Upholds standards of confidentiality	☐	☐	☐	☐
C. Demonstrates willingness to pursue professional development, training and growth opportunities	☐	☐	☐	☐
D. Observes and promotes safe work practices	☐	☐	☐	☐
E. Demonstrates flexibility in work assignments and schedule; is available for altered work schedules.	☐	☐	☐	☐
F. Demonstrates problem-solving skills and abilities	☐	☐	☐	☐
G. Adheres to and promotes RCPS School Board policies and procedures	☐	☐	☐	☐

Comments:

IV. Knowledge and performance of job responsibilities

Evaluation of Performance	Outstanding	Satisfactory	Needs Improvement	Does Not Meet
A. Maintains a consistent time of arrival at and departure from school.	☐	☐	☐	☐
B. Handles discipline problems effectively.	☐	☐	☐	☐
C. Practices emergency evacuation drills according to policy.	☐	☐	☐	☐
D. Transports only students assigned to bus unless otherwise authorized by principal with consent of parents.	☐	☐	☐	☐
E. Adheres to all state and local laws and policies.	☐	☐	☐	☐
F. Performs pre- and post-trip inspections thoroughly.	☐	☐	☐	☐

G. Demonstrates safe and defensive driving skills.	☐	☐	☐	☐
H. Maintains bus in a clean condition.	☐	☐	☐	☐
I. Submits accurate, neat reports promptly.	☐	☐	☐	☐
J. Operates school bus in accordance with route schedule and designated stops.	☐	☐	☐	☐
K. Keeps a copy of computerized bus roster in bus and adheres to.	☐	☐	☐	☐
L. Reports accidents, delays and driver absences to Transportation Office promptly.	☐	☐	☐	☐
M. Reports violators of school bus stop law promptly.	☐	☐	☐	☐
N. Reports mechanical problems to bus garage foreman promptly.	☐	☐	☐	☐

Comments:

I have read this evaluation.

Evaluator's Signature

Evaluatee's Signature

BUS DRIVER EVALUATION

Tacoma Public Schools

	Exceptional	Meets Expectations	Needs Improvement	Not Observed
Job Knowledge/Skills				
Performs pre-trip/post-trip inspection of bus to ensure it is in safe operating condition				
Keeps interior/exterior of bus clean				
Demonstrates good technical driving skills				
Is courteous, careful, attentive and alert while driving				
Skill in driving in all weather and road conditions				
Maintains appropriate schedule(s) for route(s)				
Maintains a clear and thorough route book				
Skilled in following directions and maps				
Completes Vehicle Condition Reports (VCR) as needed				
Complies with state, local and district regulations				
Assists physically handicapped children in boarding and departing from the bus as needed				
Demonstrates safe lifting procedures				
Student Management Skills				
Disciplines students using positive reinforcement				
Is cooperative, considerate, tactful and sensitive in managing students				
Deals with emergencies effectively				
Communicates effectively with teachers on discipline of students				
Personal Qualities				
Recognizes individual differences when disciplining students				
Does not lose temper or argue with students, parents, staff or co-workers				
Maintains temper and professionalism when dealing with citizens				
Is courteous to others				
Communicates positively while using the 2-way radio				
Cooperates with building administrators & transportation staff, keeping them informed of needed changes				
Adjusts easily to new assignments or changing working/driving conditions				
Reports to work on time				
Attendance				
Dress appropriately--neat, clean and well-groomed				

Name _____ Period of report _____

Date _____ Job title _____

1 - Evaluator's comments:

1. Employee's comments:

Evaluator's signature _____

Date _____

The signature below does not necessarily imply that the employee agrees with the preceding report, but only that he/she has seen and discussed it with the evaluator and/or supervisor

Employee's signature _____

Date _____

[redacted]

TERMINATION RECORD

Credit Value $2.40

1 MEMBER I.D. #

2 INDIVIDUAL'S LAST NAME ____ FIRST NAME INITIAL

3 SOCIAL SECURITY # ____ 4 PERIOD OF SERVICE: from Mo. Yr. to Mo. Yr. 5 # OF ACCIDENTS

6 STATE ____ LICENSE #

STATE ____ LICENSE #

IMPORTANT NOTICE: Refer to "Guide" for full explanation of codes below (Form G0993)

7 Eligible for Rehire (Circle only one)
001 Yes
002 Yes, but against company policy
003 No
004 Review required before rehiring

8 REASON FOR LEAVING
(Circle Only One)
101 Discharged (or Company Terminated Lease)
106 Laid Off (or Lease Suspended)
112 Leave of Absence
127 Retired
133 Resigned/Quit (or Driver Terminated Lease
199 Other____

9 STATUS
(Circle All That Apply)
202 Company Driver
207 Lease Driver (Employee of Independent Contractor)
213 Owner/Operator
228 Trip Leaser
230 Student/Trainee
234 Casual Driver
299 Other____

10 DRIVER'S EXPERIENCE
(Circle All That Apply)
303 Local
311 Mountain Driving
327 Over the Road
332 Single Driver
333 Driver Trainer
351 1st Driver of a Team
352 2nd Driver of a Team
355 Freight Handling
399 Other____

11 EQUIPMENT OPERATED
(Circle All That Apply)
504 Automobile
505 Auto Transporter
511 Bus
518 Double Trailer
523 Driveaway/Towaway
527 Dry Box
529 Dump Truck
532 Flat Bed
534 Pick Up or Hot Shot
540 Refrigerated
542 Specialized Trailer
544 Specialized Truck/Toter
547 Straight Truck
562 Tank Truck
557 Taxi
562 Triple Trailer
573 Van
581 Winch
599 Other____

12 LOADS HAULED
(Circle All That Apply)
707 Bulk Commodity
712 Container
713 Empty Trailer
714 Gen. Commodity
716 Electronics
718 Hanging Meat
720 Hazardous Material
725 Household Goods
729 Livestock
730 Lumber
731 Machinery
733 Mobile Homes
735 Motor Vehicles
762 Oversized Loads
763 Parcels
750 Passengers
764 Pipe
769 Refrigerated
773 Steel
799 Other____

13 WORK RECORD
(Circle All That Apply)
901 Satisfactory
902 Superior
903 Outstanding
911 Complaints
913 Cargo Loss
917 Equipment Loss
915 Falsified Employ. Application
924 Late Pick Up/Delivery
926 Log Violation
928 No Show
929 Failed To Report Accident
931 Quit Under Dispatch
933 Quit/Dismissed During Training and/or Orientation
957 Unauthorized Equip. Use
959 Unauthorized Passenger.
961 Unauthorized Use of Company Funds
935 Company Policy Violation
938 Unsatisfactory Safety Record
944 Personal Contact Requested
999 Other____

QUIT UNDER LOAD/ABANDONMENT
(Circle Only One)
950 Co. Terminal – With Notice
951 Auth. Location – With Notice
952 Co. Terminal – W/O Notice
953 Unauth. Location – W/O Notice
954 Left Vehicle With Team Driver
955 Unauth. Location – With Notice
956 Auth. Location – W/O Notice

14 ____
Prepared By ____ Date

15 ____
Entered By ____ Date

[redacted]

Form 0993

18 This space for changes only. TO CHANGE THIS RECORD
Record Identification Number
Signature approving change
Title

Appendix B
Alternative Termination Record Form
Work Record Section

***[redacted]* Form**
Alternative Work Record Section

	Except-ional	Meets Expect-ation	Does not meet expect-ation	Not Observed
Job Knowledge & Skills				
Performs pre/post trip truck inspections				
Demonstrates good technical driving skill				
Is courteous & attentive to other traffic				
Demonstrates skill in adverse weather				
Maintains appropriate schedules				
Maintains logs				
Maintains equipment in good working order				
Maintains cargo responsibly				
Complies with state and federal hauling regulations				
Customer Relations				
Courteous and attentive to customer concerns				
Cooperative in scheduling adjustments				
Deals with emergencies effectively				
Communicates effectively with customers				
Personal Qualities				
Recognizes individual differences among customers				
Courteous with supervisors				
Courteous with co-workers				
Responsive to directions				
Communicates effectively with supervisors				
Attendance				
Personal appearance				

OFFICE OF MANAGEMENT AND BUDGET

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies

AGENCY: Office of Management and Budget, Executive Office of the President.

ACTION: Final Guidelines.

SUMMARY: These final guidelines implement section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658). Section 515 directs the Office of Management and Budget (OMB) to issue government-wide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." By October 1, 2002, agencies must issue their own implementing guidelines that include "administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency" that does not comply with the OMB guidelines. These final guidelines also reflect the changes OMB made to the guidelines issued September 28, 2001, as a result of receiving additional comment on the "capable of being substantially reproduced" standard (paragraphs V.3.B, V.9, and V.10), which OMB previously issued on September 28, 2001, on an interim final basis.

DATES: Effective Date: January 3, 2002.

FOR FURTHER INFORMATION CONTACT: Brooke J. Dickson, Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20503. Telephone (202) 395-3785 or by e-mail to informationquality@omb.eop.gov.

SUPPLEMENTARY INFORMATION: In section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658), Congress directed the Office of Management and Budget (OMB) to issue, by September 30, 2001, government-wide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies... ." Section 515 (b) goes on to state that the OMB guidelines shall:

"(1) apply to the sharing by Federal agencies of, and access to, information disseminated by Federal agencies; and

"(2) require that each Federal agency to which the guidelines apply-

> "(A) issue guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by the agency, by

not later than 1 year after the date of issuance of the guidelines under subsection (a);

"(B) establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency that does not comply with the guidelines issued under subsection (a); and

"(C) report periodically to the Director-

"(i) the number and nature of complaints received by the agency regarding the accuracy of information disseminated by the agency; and

"(ii) how such complaints were handled by the agency."

Proposed guidelines were published in the *Federal Register* on June 28, 2001 (66 Fed. Reg. 34489). Final guidelines were published in the *Federal Register* on September 28, 2001 (66 Fed. Reg. 49718). The Supplementary Information to the final guidelines published in September 2001 provides background, the underlying principles OMB followed in issuing the final guidelines, and statements of intent concerning detailed provisions in the final guidelines.

In the final guidelines published in September 2001, OMB also requested additional comment on the "capable of being substantially reproduced" standard and the related definition of "influential scientific or statistical information" (paragraphs V.3.B, V.9, and V.10), which were issued on an interim final basis. The final guidelines published today discuss the public comments OMB received, the OMB response, and amendments to the final guidelines published in September 2001.

In developing agency-specific guidelines, agencies should refer both to the Supplementary Information to the final guidelines published in the *Federal Register* on September 28, 2001 (66 Fed. Reg. 49718), and also to the Supplementary Information published today. We stress that the three "Underlying Principles" that OMB followed in drafting the guidelines that we published on September 28, 2001 (66 Fed. Reg. 49719), are also applicable to the amended guidelines that we publish today.

In accordance with section 515, OMB has designed the guidelines to help agencies ensure and maximize the quality, utility, objectivity and integrity of the information that they disseminate (meaning to share with, or give access to, the public). It is crucial that information Federal agencies disseminate meets these guidelines. In this respect, the fact that the Internet enables agencies to communicate information quickly and easily to a wide audience not only offers great benefits to society, but also increases the potential harm that can result from the dissemination of information that does not meet basic information quality guidelines. Recognizing the wide variety of information Federal agencies disseminate and the wide variety of dissemination practices that agencies have, OMB developed the guidelines with several principles in mind.

First, OMB designed the guidelines to apply to a wide variety of government information dissemination activities that may range in importance and scope. OMB also designed the guidelines to be generic enough to fit all media, be they printed, electronic, or in other form. OMB sought to avoid the problems that would be inherent in developing detailed, prescriptive, "one-size-fits-all" government-wide guidelines that would artificially require different types of

dissemination activities to be treated in the same manner. Through this flexibility, each agency will be able to incorporate the requirements of these OMB guidelines into the agency's own information resource management and administrative practices.

Second, OMB designed the guidelines so that agencies will meet basic information quality standards. Given the administrative mechanisms required by section 515 as well as the standards set forth in the Paperwork Reduction Act, it is clear that agencies should not disseminate substantive information that does not meet a basic level of quality. We recognize that some government information may need to meet higher or more specific information quality standards than those that would apply to other types of government information. The more important the information, the higher the quality standards to which it should be held, for example, in those situations involving "influential scientific, financial, or statistical information" (a phrase defined in these guidelines). The guidelines recognize, however, that information quality comes at a cost. Accordingly, the agencies should weigh the costs (for example, including costs attributable to agency processing effort, respondent burden, maintenance of needed privacy, and assurances of suitable confidentiality) and the benefits of higher information quality in the development of information, and the level of quality to which the information disseminated will be held.

Third, OMB designed the guidelines so that agencies can apply them in a common-sense and workable manner. It is important that these guidelines do not impose unnecessary administrative burdens that would inhibit agencies from continuing to take advantage of the Internet and other technologies to disseminate information that can be of great benefit and value to the public. In this regard, OMB encourages agencies to incorporate the standards and procedures required by these guidelines into their existing information resources management and administrative practices rather than create new and potentially duplicative or contradictory processes. The primary example of this is that the guidelines recognize that, in accordance with OMB Circular A-130, agencies already have in place well-established information quality standards and administrative mechanisms that allow persons to seek and obtain correction of information that is maintained and disseminated by the agency. Under the OMB guidelines, agencies need only ensure that their own guidelines are consistent with these OMB guidelines, and then ensure that their administrative mechanisms satisfy the standards and procedural requirements in the new agency guidelines. Similarly, agencies may rely on their implementation of the Federal Government's computer security laws (formerly, the Computer Security Act, and now the computer security provisions of the Paperwork Reduction Act) to establish appropriate security safeguards for ensuring the "integrity" of the information that the agencies disseminate.

In addition, in response to concerns expressed by some of the agencies, we want to emphasize that OMB recognizes that Federal agencies provide a wide variety of data and information. Accordingly, OMB understands that the guidelines discussed below cannot be implemented in the same way by each agency. In some cases, for example, the data disseminated by an agency are not collected by that agency; rather, the information the agency must provide in a timely manner is compiled from a variety of sources that are constantly updated and revised and may be confidential. In such cases, while agencies' implementation of the guidelines may differ, the essence of the guidelines will apply. That is, these agencies must make their methods transparent by providing documentation, ensure quality by reviewing the underlying methods used in developing the data and consulting (as appropriate) with experts and users, and keep users informed about corrections and revisions.

Summary of OMB Guidelines

These guidelines apply to Federal agencies subject to the Paperwork Reduction Act (44 U.S.C. chapter 35). Agencies are directed to develop information resources management procedures for reviewing and substantiating (by documentation or other means selected by the agency) the quality (including the objectivity, utility, and integrity) of information before it is disseminated. In addition, agencies are to establish administrative mechanisms allowing affected persons to seek and obtain, where appropriate, correction of information disseminated by the agency that does not comply with the OMB or agency guidelines. Consistent with the underlying principles described above, these guidelines stress the importance of having agencies apply these standards and develop their administrative mechanisms so they can be implemented in a common sense and workable manner. Moreover, agencies must apply these standards flexibly, and in a manner appropriate to the nature and timeliness of the information to be disseminated, and incorporate them into existing agency information resources management and administrative practices.

Section 515 denotes four substantive terms regarding information disseminated by Federal agencies: quality, utility, objectivity, and integrity. It is not always clear how each substantive term relates -- or how the four terms in aggregate relate -- to the widely divergent types of information that agencies disseminate. The guidelines provide definitions that attempt to establish a clear meaning so that both the agency and the public can readily judge whether a particular type of information to be disseminated does or does not meet these attributes.

In the guidelines, OMB defines "quality" as the encompassing term, of which "utility," "objectivity," and "integrity" are the constituents. "Utility" refers to the usefulness of the information to the intended users. "Objectivity" focuses on whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner, and as a matter of substance, is accurate, reliable, and unbiased. "Integrity" refers to security -- the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification. OMB modeled the definitions of "information," "government information," "information dissemination product," and "dissemination" on the longstanding definitions of those terms in OMB Circular A-130, but tailored them to fit into the context of these guidelines.

In addition, Section 515 imposes two reporting requirements on the agencies. The first report, to be promulgated no later than October 1, 2002, must provide the agency's information quality guidelines that describe administrative mechanisms allowing affected persons to seek and obtain, where appropriate, correction of disseminated information that does not comply with the OMB and agency guidelines. The second report is an annual fiscal year report to OMB (to be first submitted on January 1, 2004) providing information (both quantitative and qualitative, where appropriate) on the number, nature, and resolution of complaints received by the agency regarding its perceived or confirmed failure to comply with these OMB and agency guidelines.

Public Comments and OMB Response

Applicability of Guidelines. Some comments raised concerns about the applicability of these guidelines, particularly in the context of scientific research conducted by Federally employed scientists or Federal grantees who publish and communicate their research findings in the same manner as their academic colleagues. OMB believes that information generated and disseminated in these contexts is not covered by these guidelines unless the agency represents the information as, or uses the information in support of, an official position of the agency.

As a general matter, these guidelines apply to "information" that is "disseminated" by agencies subject to the Paperwork Reduction Act (44 U.S.C. 3502(1)). See paragraphs II, V.5 and V.8. The definitions of "information" and "dissemination" establish the scope of the applicability of these guidelines. "Information" means "any communication or representation of knowledge such as facts or data ..." This definition of information in paragraph V.5 does "not include opinions, where the agency's presentation makes it clear that what is being offered is someone's opinion rather than fact or the agency's views."

"Dissemination" is defined to mean "agency initiated or sponsored distribution of information to the public." As used in paragraph V.8, "agency INITIATED ... distribution of information to the public" refers to information that the agency disseminates, e.g., a risk assessment prepared by the agency to inform the agency's formulation of possible regulatory or other action. In addition, if an agency, as an institution, disseminates information prepared by an outside party in a manner that reasonably suggests that the agency agrees with the information, this appearance of having the information represent agency views makes agency dissemination of the information subject to these guidelines. By contrast, an agency does not "initiate" the dissemination of information when a Federally employed scientist or Federal grantee or contractor publishes and communicates his or her research findings in the same manner as his or her academic colleagues, even if the Federal agency retains ownership or other intellectual property rights because the Federal government paid for the research. To avoid confusion regarding whether the agency agrees with the information (and is therefore disseminating it through the employee or grantee), the researcher should include an appropriate disclaimer in the publication or speech to the effect that the "views are mine, and do not necessarily reflect the view" of the agency.

Similarly, as used in paragraph V.8, "agency ... SPONSORED distribution of information to the public" refers to situations where an agency has directed a third-party to disseminate information, or where the agency has the authority to review and approve the information before release. Therefore, for example, if an agency through a procurement contract or a grant provides for a person to conduct research, and then the agency directs the person to disseminate the results (or the agency reviews and approves the results before they may be disseminated), then the agency has "sponsored" the dissemination of this information. By contrast, if the agency simply provides funding to support research, and it the researcher (not the agency) who decides whether to disseminate the results and - if the results are to be released - who determines the content and presentation of the dissemination, then the agency has not "sponsored" the dissemination even though it has funded the research and even if the Federal agency retains ownership or other intellectual property rights because the Federal government paid for the research. To avoid confusion regarding whether the agency is sponsoring the dissemination, the researcher should include an appropriate disclaimer in the publication or speech to the effect that the "views are mine, and do not necessarily reflect the view" of the agency. On the other hand, subsequent agency dissemination of such information requires that the information adhere to the agency's information quality guidelines. In sum, these guidelines govern an agency's dissemination of information, but generally do not govern a third-party's dissemination of information (the exception being where the agency is essentially using the third-party to disseminate information on the agency's behalf). Agencies, particularly those that fund scientific research, are encouraged to clarify the applicability of these guidelines to the various types of information they and their employees and grantees disseminate.

Paragraph V.8 also states that the definition of "dissemination" does not include "...

distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes." The exemption from the definition of "dissemination" for "adjudicative processes" is intended to exclude, from the scope of these guidelines, the findings and determinations that an agency makes in the course of adjudications involving specific parties. There are well-established procedural safeguards and rights to address the quality of adjudicatory decisions and to provide persons with an opportunity to contest decisions. These guidelines do not impose any additional requirements on agencies during adjudicative proceedings and do not provide parties to such adjudicative proceedings any additional rights of challenge or appeal.

The Presumption Favoring Peer-Reviewed Information. As a general matter, in the scientific and research context, we regard technical information that has been subjected to formal, independent, external peer review as presumptively objective. As the guidelines state in paragraph V.3.b.i: "If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity." An example of a formal, independent, external peer review is the review process used by scientific journals.

Most comments approved of the prominent role that peer review plays in the OMB guidelines. Some comments contended that peer review was not accepted as a universal standard that incorporates an established, practiced, and sufficient level of objectivity. Other comments stated that the guidelines would be better clarified by making peer review one of several factors that an agency should consider in assessing the objectivity (and quality in general) of original research. In addition, several comments noted that peer review does not establish whether analytic results are capable of being substantially reproduced. In light of the comments, the final guidelines in new paragraph V.3.b.i qualify the presumption in favor of peer-reviewed information as follows: "However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance."

We believe that transparency is important for peer review, and these guidelines set minimum standards for the transparency of agency-sponsored peer review. As we state in new paragraph V.3.b.i: "If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity. However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance. If agency-sponsored peer review is employed to help satisfy the objectivity standard, the review process employed shall meet the general criteria for competent and credible peer review recommended by OMB-OIRA to the President's Management Council (9/20/01) (http://www.whitehouse.gov/omb/inforeg/oira_review-process.html), namely, 'that (a) peer reviewers be selected primarily on the basis of necessary technical expertise, (b) peer reviewers be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers be expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and (d) peer reviews be conducted in an open and rigorous manner.' "

The importance of these general criteria for competent and credible peer review has been supported by a number of expert bodies. For example, "the work of fully competent peer-review panels can be undermined by allegations of conflict of interest and bias. Therefore, the best interests of the Board are served by effective policies and procedures regarding potential conflicts of interest, impartiality, and panel balance." (*EPA's Science Advisory Board Panels:*

Improved Policies and Procedures Needed to Ensure Independence and Balance," GAO-01-536, General Accounting Office, Washington, D.C., June 2001, page 19.) As another example, "risk analyses should be peer-reviewed and accessible - both physically and intellectually - so that decision-makers at all levels will be able to respond critically to risk characterizations. The intensity of the peer reviews should be commensurate with the significance of the risk or its management implications." (*Setting Priorities, Getting Results: A New Direction for EPA*, Summary Report, National Academy of Public Administration, Washington, D.C. April, 1995, page 23.)

These criteria for peer reviewers are generally consistent with the practices now followed by the National Research Council of the National Academy of Sciences. In considering these criteria for peer reviewers, we note that there are many types of peer reviews and that agency guidelines concerning the use of peer review should tailor the rigor of peer review to the importance of the information involved. More generally, agencies should define their peer-review standards in appropriate ways, given the nature and importance of the information they disseminate.

Is Journal Peer Review Always Sufficient? Some comments argued that journal peer review should be adequate to demonstrate quality, even for influential information that can be expected to have major effects on public policy. OMB believes that this position overstates the effectiveness of journal peer review as a quality-control mechanism.

Although journal peer review is clearly valuable, there are cases where flawed science has been published in respected journals. For example, the NIH Office of Research Integrity recently reported the following case regarding environmental health research:

"Based on the report of an investigation conducted by [XX] University, dated July 16, 1999, and additional analysis conducted by ORI in its oversight review, the US Public Health Service found that Dr. [X] engaged in scientific misconduct. Dr. [X] committed scientific misconduct by intentionally falsifying the research results published in the journal SCIENCE and by providing falsified and fabricated materials to investigating officials at [XX] University in response to a request for original data to support the research results and conclusions report in the SCIENCE paper. In addition, PHS finds that there is no original data or other corroborating evidence to support the research results and conclusions reported in the SCIENCE paper as a whole." (66 Fed. Reg. 52137, October 12, 2001).

Although such cases of falsification are presumably rare, there is a significant scholarly literature documenting quality problems with articles published in peer-reviewed research. "In a [peer-reviewed] meta-analysis that surprised many -- and some doubt -- researchers found little evidence that peer review actually improves the quality of research papers." (See, e.g., *Science*, Vol. 293, page 2187 (September 21, 2001.)) In part for this reason, many agencies have already adopted peer review and science advisory practices that go beyond journal peer review. See, e.g., Sheila Jasanoff, *The Fifth Branch: Science Advisers as Policy Makers*, Cambridge, MA, Harvard University Press, 1990; Mark R. Powell, *Science at EPA: Information in the Regulatory Process*, Resources for the Future, Washington, D.C., 1999, pages 138-139; 151-153; *Implementation of the Environmental Protection Agency's Peer Review Program: An SAB Evaluation of Three Reviews*, EPA-SAB-RSAC-01-009, A Review of the Research Strategies Advisory Committee (RSAC) of the EPA Science Advisory Board (SAB), Washington, D.C., September 26, 2001. For information likely to have an important public policy or private sector impact, OMB believes

that additional quality checks beyond peer review are appropriate.

Definition of "Influential". OMB guidelines apply stricter quality standards to the dissemination of information that is considered "influential." Comments noted that the breadth of the definition of "influential" in interim final paragraph V.9 requires much speculation on the part of agencies.

We believe that this criticism has merit and have therefore narrowed the definition. In this narrower definition, "influential", when used in the phrase "influential scientific, financial, or statistical information", is amended to mean that "the agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions." The intent of the new phrase "clear and substantial" is to reduce the need for speculation on the part of agencies. We added the present tense -- "or does have" -- to this narrower definition because on occasion, an information dissemination may occur simultaneously with a particular policy change. In response to a public comment, we added an explicit reference to "financial" information as consistent with our original intent.

Given the differences in the many Federal agencies covered by these guidelines, and the differences in the nature of the information they disseminate, we also believe it will be helpful if agencies elaborate on this definition of "influential" in the context of their missions and duties, with due consideration of the nature of the information they disseminate. As we state in amended paragraph V.9, "Each agency is authorized to define 'influential' in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible."

Reproducibility. As we state in new paragraph V.3.b.ii: " If an agency is responsible for disseminating influential scientific, financial, or statistical information, agency guidelines shall include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties." OMB believes that a reproducibility standard is practical and appropriate for information that is considered "influential", as defined in paragraph V.9 - that "will have or does have a clear and substantial impact on important public policies or important private sector decisions." The reproducibility standard applicable to influential scientific, financial, or statistical information is intended to ensure that information disseminated by agencies is sufficiently transparent in terms of data and methods of analysis that it would be feasible for a replication to be conducted. The fact that the use of original and supporting data and analytic results have been deemed "defensible" by peer-review procedures does not necessarily imply that the results are transparent and replicable.

Reproducibility of Original and Supporting Data. Several of the comments objected to the exclusion of original and supporting data from the reproducibility requirements. Comments instead suggested that OMB should apply the reproducibility standard to original data, and that OMB should provide flexibility to the agencies in determining what constitutes "original and supporting" data. OMB agrees and asks that agencies consider, in developing their own guidelines, which categories of original and supporting data should be subject to the reproducibility standard and which should not. To help in resolving this issue, we also ask agencies to consult directly with relevant scientific and technical communities on the feasibility of having the selected categories of original and supporting data subject to the reproducibility standard. Agencies are encouraged to address ethical, feasibility, and confidentiality issues with care. As we state in new paragraph V.3.b.ii.A, "Agencies may identify, in consultation with the

relevant scientific and technical communities, those particular types of data that can practicably be subjected to a reproducibility requirement, given ethical, feasibility, or confidentiality constraints." Further, as we state in our expanded definition of "reproducibility" in paragraph V.10, "If agencies apply the reproducibility test to specific types of original or supporting data, the associated guidelines shall provide relevant definitions of reproducibility (e.g., standards for replication of laboratory data)." OMB urges caution in the treatment of original and supporting data because it may often be impractical or even impermissible or unethical to apply the reproducibility standard to such data. For example, if may not be ethical to repeat a "negative" (ineffective) clinical (therapeutic) experiment and it may not be feasible to replicate the radiation exposures studied after the Chernobyl accident. When agencies submit their draft agency guidelines for OMB review, agencies should include a description of the extent to which the reproducibility standard is applicable and reflect consultations with relevant scientific and technical communities that were used in developing guidelines related to applicability of the reproducibility standard to original and supporting data.

It is also important to emphasize that the reproducibility standard does not apply to all original and supporting data disseminated by agencies. As we state in new paragraph V.3.b.ii.A, "With regard to original and supporting data related [to influential scientific, financial, or statistical information], agency guidelines shall not require that all disseminated data be subjected to a reproducibility requirement." In addition, we encourage agencies to address how greater transparency can be achieved regarding original and supporting data. As we also state in new paragraph V.3.b.ii.A, "It is understood that reproducibility of data is an indication of transparency about research design and methods and thus a replication exercise (i.e., a new experiment, test, or sample) shall not be required prior to each dissemination." Agency guidelines need to achieve a high degree of transparency about data even when reproducibility is not required.

Reproducibility of Analytic Results. Many public comments were critical of the reproducibility standard and expressed concern that agencies would be required to reproduce each analytic result before it is disseminated. While several comments commended OMB for establishing an appropriate balance in the "capable of being substantially reproduced" standard, others considered this standard to be inherently subjective. There were also comments that suggested the standard would cause more burden for agencies.

It is not OMB's intent that each agency must reproduce each analytic result before it is disseminated. The purpose of the reproducibility standard is to cultivate a consistent agency commitment to transparency about how analytic results are generated: the specific data used, the various assumptions employed, the specific analytic methods applied, and the statistical procedures employed. If sufficient transparency is achieved on each of these matters, then an analytic result should meet the "capable of being substantially reproduced" standard.

While there is much variation in types of analytic results, OMB believes that reproducibility is a practical standard to apply to most types of analytic results. As we state in new paragraph V.3.b.ii.B, "With regard to analytic results related [to influential scientific, financial, or statistical information], agency guidelines shall generally require sufficient transparency about data and methods that an independent reanalysis could be undertaken by a qualified member of the public. These transparency standards apply to agency analysis of data from a single study as well as to analyses that combine information from multiple studies." We elaborate upon this principle in our expanded definition of "reproducibility" in paragraph V.10:

"With respect to analytic results, 'capable of being substantially reproduced' means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error."

Even in a situation where the original and supporting data are protected by confidentiality concerns, or the analytic computer models or other research methods may be kept confidential to protect intellectual property, it may still be feasible to have the analytic results subject to the reproducibility standard. For example, a qualified party, operating under the same confidentiality protections as the original analysts, may be asked to use the same data, computer model or statistical methods to replicate the analytic results reported in the original study. See, e.g., "Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality," A Special Report of the Health Effects Institute's Particle Epidemiology Reanalysis Project, Cambridge, MA, 2000.

The primary benefit of public transparency is not necessarily that errors in analytic results will be detected, although error correction is clearly valuable. The more important benefit of transparency is that the public will be able to assess how much an agency's analytic result hinges on the specific analytic choices made by the agency. Concreteness about analytic choices allows, for example, the implications of alternative technical choices to be readily assessed. This type of sensitivity analysis is widely regarded as an essential feature of high-quality analysis, yet sensitivity analysis cannot be undertaken by outside parties unless a high degree of transparency is achieved. The OMB guidelines do not compel such sensitivity analysis as a necessary dimension of quality, but the transparency achieved by reproducibility will allow the public to undertake sensitivity studies of interest.

We acknowledge that confidentiality concerns will sometimes preclude public access as an approach to reproducibility. In response to public comment, we have clarified that such concerns do include interests in "intellectual property." To ensure that the OMB guidelines have sufficient flexibility with regard to analytic transparency, OMB has, in new paragraph V.3.b.ii.B.*i*, provided agencies an alternative approach for classes or types of analytic results that cannot practically be subject to the reproducibility standard. "[In those situations involving influential scientific, financial, or statistical information ...] making the data and methods publicly available will assist in determining whether analytic results are reproducible. However, the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections." Specifically, in cases where reproducibility will not occur due to other compelling interests, we expect agencies (1) to perform robustness checks appropriate to the importance of the information involved, e.g., determining whether a specific statistic is sensitive to the choice of analytic method, and, accompanying the information disseminated, to document their efforts to assure the needed robustness in information quality, and (2) address in their guidelines the degree to which they anticipate the opportunity for reproducibility to be limited by the confidentiality of underlying data. ii, "In situations where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken. Agency guidelines shall, however, in all cases, require a disclosure of the specific data sources that have been used and the specific quantitative methods and assumptions that have been employed."

Given the differences in the many Federal agencies covered by these guidelines, and the differences in robustness checks and the level of detail for documentation thereof that might be

appropriate for different agencies, we also believe it will be helpful if agencies elaborate on these matters in the context of their missions and duties, with due consideration of the nature of the information they disseminate. As we state in new paragraph V.3.b.ii.B.*ii*, "Each agency is authorized to define the type of robustness checks, and the level of detail for documentation thereof, in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.

We leave the determination of the appropriate degree of rigor to the discretion of agencies and the relevant scientific and technical communities that work with the agencies. We do, however, establish a general standard for the appropriate degree of rigor in our expanded definition of "reproducibility" in paragraph V.10: "'Reproducibility' means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased)." OMB will review each agency's treatment of this issue when reviewing the agency guidelines as a whole.

Comments also expressed concerns regarding interim final paragraph V.3.B.iii, "making the data and models publicly available will assist in determining whether analytic results are capable of being substantially reproduced," and whether it could be interpreted to constitute public dissemination of these materials, rendering moot the reproducibility test. (For the equivalent provision, see new paragraph V.3.b.ii.B.*i*.) The OMB guidelines do not require agencies to reproduce each disseminated analytic result by independent reanalysis. Thus, public dissemination of data and models *per se* does not mean that the analytic result has been reproduced. It means only that the result should be CAPABLE of being reproduced. The transparency associated with this capability of reproduction is what the OMB guidelines are designed to achieve.

We also want to build on a general observation that we made in our final guidelines published in September 2001. In those guidelines we stated: "... in those situations involving influential scientific[, financial,] or statistical information, the substantial reproducibility standard is added as a quality standard above and beyond some peer review quality standards" (66 Fed. Reg. 49722 (September 28, 2001)). A hypothetical example may serve to illustrate this point. Assume that two Federal agencies initiated or sponsored the dissemination of five scientific studies after October 1, 2002 (see paragraph III.4) that were, before dissemination, subjected to formal, independent, external peer review, i.e., that met the presumptive standard for "objectivity" under paragraph V.3.b.i. Further assume, at the time of dissemination, that neither agency reasonably expected that the dissemination of any of these studies would have "a clear and substantial impact" on important public policies, i.e., that these studies were not considered "influential" under paragraph V.9, and thus not subject to the reproducibility standards in paragraphs V.3.b.ii.A or B. Then assume, two years later, in 2005, that one of the agencies decides to issue an important and far-reaching regulation based clearly and substantially on the agency's evaluation of the analytic results set forth in these five studies and that such agency reliance on these five studies as published in the agency's notice of proposed rulemaking would constitute dissemination of these five studies. These guidelines would require the rulemaking agency, prior to publishing the notice of proposed rulemaking, to evaluate these five studies to determine if the analytic results stated therein would meet the "capable of being substantially reproduced" standards in paragraph V.3.b.ii.B and, if necessary, related standards governing original and supporting data in paragraph V.3.b.ii.A. If the agency were to decide that any of the five studies would not meet the reproducibility standard, the agency may still rely on them but

only if they satisfy the transparency standard and - as applicable - the disclosure of robustness checks required by these guidelines. Otherwise, the agency should not disseminate any of the studies that did not meet the applicable standards in the guidelines at the time it publishes the notice of proposed rulemaking.

Some comments suggested that OMB consider replacing the reproducibility standard with a standard concerning "confirmation" of results for influential scientific and statistical information. Although we encourage agencies to consider "confirmation" as a relevant standard -- at least in some cases -- for assessing the objectivity of original and supporting data, we believe that "confirmation" is too stringent a standard to apply to analytic results. Often the regulatory impact analysis prepared by an agency for a major rule, for example, will be the only formal analysis of an important subject. It would be unlikely that the results of the regulatory impact analysis had already been confirmed by other analyses. The "capable of being substantially reproduced" standard is less stringent than a "confirmation" standard because it simply requires that an agency's analysis be sufficiently transparent that another qualified party could replicate it through reanalysis.

Health, Safety, and Environmental Information. We note, in the scientific context, that in 1996 the Congress, for health decisions under the Safe Drinking Water Act, adopted a basic standard of quality for the use of science in agency decisionmaking. Under 42 U.S.C. 300g-1(b)(3)(A), an agency is directed, "to the degree that an Agency action is based on science," to use "(i) the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices; and (ii) data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data)."

We further note that in the 1996 amendments to the Safe Drinking Water Act, Congress adopted a basic quality standard for the dissemination of public information about risks of adverse health effects. Under 42 U.S.C. 300g-1(b)(3)(B), the agency is directed, "to ensure that the presentation of information [risk] effects is comprehensive, informative, and understandable." The agency is further directed, "in a document made available to the public in support of a regulation [to] specify, to the extent practicable - (i) each population addressed by any estimate [of applicable risk effects]; (ii) the expected risk or central estimate of risk for the specific populations [affected]; (iii) each appropriate upper-bound or lower-bound estimate of risk; (iv) each significant uncertainty identified in the process of the assessment of [risk] effects and the studies that would assist in resolving the uncertainty; and (v) peer-reviewed studies known to the [agency] that support, are directly relevant to, or fail to support any estimate of [risk] effects and the methodology used to reconcile inconsistencies in the scientific data."

As suggested in several comments, we have included these congressional standards directly in new paragraph V.3.b.ii.C, and made them applicable to the information disseminated by all the agencies subject to these guidelines: "With regard to analysis of risks to human health, safety and the environment maintained or disseminated by the agencies, agencies shall either adopt or adapt the quality principles applied by Congress to risk information used and disseminated pursuant to the Safe Drinking Water Act Amendments of 1996 (42 U.S.C. 300g-1 (b)(3)(A) & (B))." The word "adapt" is intended to provide agencies flexibility in applying these principles to various types of risk assessment.

Comments also argued that the continued flow of vital information from agencies

responsible for disseminating health and medical information to medical providers, patients, and the public may be disrupted due to these peer review and reproducibility standards. OMB responded by adding to new paragraph V.3.b.ii.C: "Agencies responsible for dissemination of vital health and medical information shall interpret the reproducibility and peer-review standards in a manner appropriate to assuring the timely flow of vital information from agencies to medical providers, patients, health agencies, and the public. Information quality standards may be waived temporarily by agencies under urgent situations (e.g., imminent threats to public health or homeland security) in accordance with the latitude specified in agency-specific guidelines."

Administrative Correction Mechanisms. In addition to commenting on the substantive standards in these guidelines, many of the comments noted that the OMB guidelines on the administrative correction of information do not specify a time period in which the agency investigation and response must be made. OMB has added the following new paragraph III.3.i to direct agencies to specify appropriate time periods in which the investigation and response need to be made. "Agencies shall specify appropriate time periods for agency decisions on whether and how to correct the information, and agencies shall notify the affected persons of the corrections made."

Several comments stated that the OMB guidelines needed to direct agencies to consider incorporating an administrative appeal process into their administrative mechanisms for the correction of information. OMB agreed, and added the following new paragraph III.3.ii: "If the person who requested the correction does not agree with the agency's decision (including the corrective action, if any), the person may file for reconsideration within the agency. The agency shall establish an administrative appeal process to review the agency's initial decision, and specify appropriate time limits in which to resolve such requests for reconsideration." Recognizing that many agencies already have a process in place to respond to public concerns, it is not necessarily OMB's intent to require these agencies to establish a new or different process. Rather, our intent is to ensure that agency guidelines specify an objective administrative appeal process that, upon further complaint by the affected person, reviews an agency's decision to disagree with the correction request. An objective process will ensure that the office that originally disseminates the information does not have responsibility for both the initial response and resolution of a disagreement. In addition, the agency guidelines should specify that if the agency believes other agencies may have an interest in the resolution of any administrative appeal, the agency should consult with those other agencies about their possible interest.

Overall, OMB does not envision administrative mechanisms that would burden agencies with frivolous claims. Instead, the correction process should serve to address the genuine and valid needs of the agency and its constituents without disrupting agency processes. Agencies, in making their determination of whether or not to correct information, may reject claims made in bad faith or without justification, and are required to undertake only the degree of correction that they conclude is appropriate for the nature and timeliness of the information involved, and explain such practices in their annual fiscal year reports to OMB.

OMB's issuance of these final guidelines is the beginning of an evolutionary process that will include draft agency guidelines, public comment, final agency guidelines, development of experience with OMB and agency guidelines, and continued refinement of both OMB and agency guidelines. Just as OMB requested public comment before issuing these final guidelines, OMB will refine these guidelines as experience develops and further public comment is obtained.

Dated: December 21, 2001

/s/
John D. Graham, Ph.D.
Administrator, Office of Information and Regulatory Affairs

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies.

I. OMB Responsibilities. Section 515 of the Treasury and General Government Appropriations Act for FY2001 (Public Law 106-554) directs the Office of Management and Budget to issue government-wide guidelines that provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by Federal agencies.

II. Agency Responsibilities. Section 515 directs agencies subject to the Paperwork Reduction Act (44 U.S.C. 3502(1)) to -

1. Issue their own information quality guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by the agency no later than one year after the date of issuance of the OMB guidelines;

2. Establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency that does not comply with these OMB guidelines; and

3. Report to the Director of OMB the number and nature of complaints received by the agency regarding agency compliance with these OMB guidelines concerning the quality, objectivity, utility, and integrity of information and how such complaints were resolved.

III. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies.

1. Overall, agencies shall adopt a basic standard of quality (including objectivity, utility, and integrity) as a performance goal and should take appropriate steps to incorporate information quality criteria into agency information dissemination practices. Quality is to be ensured and established at levels appropriate to the nature and timeliness of the information to be disseminated. Agencies shall adopt specific standards of quality that are appropriate for the various categories of information they disseminate.

2. As a matter of good and effective agency information resources management, agencies shall develop a process for reviewing the quality (including the objectivity, utility, and integrity) of information before it is disseminated. Agencies shall treat information quality as integral to every step of an agency's development of information, including creation, collection, maintenance, and dissemination. This process shall enable the agency to substantiate the quality of the information it has disseminated through documentation or other means appropriate to the information.

3. To facilitate public review, agencies shall establish administrative mechanisms allowing affected persons to seek and obtain, where appropriate, timely correction of information maintained and disseminated by the agency that does not comply with OMB or agency guidelines. These administrative mechanisms shall be flexible, appropriate to the nature and timeliness of the disseminated information, and incorporated into agency information resources management and administrative practices.

    i. Agencies shall specify appropriate time periods for agency decisions on whether and how to correct the information, and agencies shall notify the affected persons of the corrections made.

    ii. If the person who requested the correction does not agree with the agency's decision (including the corrective action, if any), the person may file for reconsideration within the agency. The agency shall establish an administrative appeal process to review the agency's initial decision, and specify appropriate time limits in which to resolve such requests for reconsideration.

4. The agency's pre-dissemination review, under paragraph III.2, shall apply to information that the agency first disseminates on or after October 1, 2002. The agency's administrative mechanisms, under paragraph III.3, shall apply to information that the agency disseminates on or after October 1, 2002, regardless of when the agency first disseminated the information.

IV. Agency Reporting Requirements.

1. Agencies must designate the Chief Information Officer or another official to be responsible for agency compliance with these guidelines.

2. The agency shall respond to complaints in a manner appropriate to the nature and extent of the complaint. Examples of appropriate responses include personal contacts via letter or telephone, form letters, press releases or mass mailings that correct a widely disseminated error or address a frequently raised complaint.

3. Each agency must prepare a draft report, no later than April 1, 2002, providing the agency's information quality guidelines and explaining how such guidelines will ensure and maximize the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by the agency. This report must also detail the administrative mechanisms developed by that agency to allow affected persons to seek and obtain appropriate correction of information maintained and disseminated by the agency that does not comply with the OMB or the agency guidelines.

4. The agency must publish a notice of availability of this draft report in the *Federal Register*, and post this report on the agency's website, to provide an opportunity for public comment.

5. Upon consideration of public comment and after appropriate revision, the agency must submit this draft report to OMB for review regarding consistency with these OMB guidelines no later than July 1, 2002. Upon completion of that OMB review and completion of this report, agencies must publish notice of the availability of this report in

its final form in the *Federal Register*, and post this report on the agency's web site no later than October 1, 2002.

6. On an annual fiscal-year basis, each agency must submit a report to the Director of OMB providing information (both quantitative and qualitative, where appropriate) on the number and nature of complaints received by the agency regarding agency compliance with these OMB guidelines and how such complaints were resolved. Agencies must submit these reports no later than January 1 of each following year, with the first report due January 1, 2004.

V. Definitions.

1. "Quality" is an encompassing term comprising utility, objectivity, and integrity. Therefore, the guidelines sometimes refer to these four statutory terms, collectively, as "quality."

2. "Utility" refers to the usefulness of the information to its intended users, including the public. In assessing the usefulness of information that the agency disseminates to the public, the agency needs to consider the uses of the information not only from the perspective of the agency but also from the perspective of the public. As a result, when transparency of information is relevant for assessing the information's usefulness from the public's perspective, the agency must take care to ensure that transparency has been addressed in its review of the information.

3. "Objectivity" involves two distinct elements, presentation and substance.

    a. "Objectivity" includes whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner. This involves whether the information is presented within a proper context. Sometimes, in disseminating certain types of information to the public, other information must also be disseminated in order to ensure an accurate, clear, complete, and unbiased presentation. Also, the agency needs to identify the sources of the disseminated information (to the extent possible, consistent with confidentiality protections) and, in a scientific, financial, or statistical context, the supporting data and models, so that the public can assess for itself whether there may be some reason to question the objectivity of the sources. Where appropriate, data should have full, accurate, transparent documentation, and error sources affecting data quality should be identified and disclosed to users.

    b. In addition, "objectivity" involves a focus on ensuring accurate, reliable, and unbiased information. In a scientific, financial, or statistical context, the original and supporting data shall be generated, and the analytic results shall be developed, using sound statistical and research methods.

        i. If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity. However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance. If agency-sponsored peer review is employed to help satisfy the objectivity standard, the review process employed shall meet the general criteria for competent and

credible peer review recommended by OMB-OIRA to the President's Management Council (9/20/01) (http://www.whitehouse.gov/omb/inforeg/oira_review-process.html), namely, "that (a) peer reviewers be selected primarily on the basis of necessary technical expertise, (b) peer reviewers be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers be expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and (d) peer reviews be conducted in an open and rigorous manner."

ii. If an agency is responsible for disseminating influential scientific, financial, or statistical information, agency guidelines shall include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties.

A. With regard to original and supporting data related thereto, agency guidelines shall not require that all disseminated data be subjected to a reproducibility requirement. Agencies may identify, in consultation with the relevant scientific and technical communities, those particular types of data that can practicably be subjected to a reproducibility requirement, given ethical, feasibility, or confidentiality constraints. It is understood that reproducibility of data is an indication of transparency about research design and methods and thus a replication exercise (i.e., a new experiment, test, or sample) shall not be required prior to each dissemination.

B. With regard to analytic results related thereto, agency guidelines shall generally require sufficient transparency about data and methods that an independent reanalysis could be undertaken by a qualified member of the public. These transparency standards apply to agency analysis of data from a single study as well as to analyses that combine information from multiple studies.

i. Making the data and methods publicly available will assist in determining whether analytic results are reproducible. However, the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections.

ii. In situations where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken. Agency guidelines shall, however, in all cases, require a disclosure of the specific data sources that have been used and the specific quantitative methods and assumptions that have been employed. Each agency is authorized to define the type of robustness checks, and the level of detail for documentation thereof, in ways appropriate for it given the nature and multiplicity of issues for which the agency is

responsible.

C. With regard to analysis of risks to human health, safety and the environment maintained or disseminated by the agencies, agencies shall either adopt or adapt the quality principles applied by Congress to risk information used and disseminated pursuant to the Safe Drinking Water Act Amendments of 1996 (42 U.S.C. 300g-1(b)(3)(A) & (B)). Agencies responsible for dissemination of vital health and medical information shall interpret the reproducibility and peer-review standards in a manner appropriate to assuring the timely flow of vital information from agencies to medical providers, patients, health agencies, and the public. Information quality standards may be waived temporarily by agencies under urgent situations (e.g., imminent threats to public health or homeland security) in accordance with the latitude specified in agency-specific guidelines.

4. "Integrity" refers to the security of information -- protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.

5. "Information" means any communication or representation of knowledge such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms. This definition includes information that an agency disseminates from a web page, but does not include the provision of hyperlinks to information that others disseminate. This definition does not include opinions, where the agency's presentation makes it clear that what is being offered is someone's opinion rather than fact or the agency's views.

6. "Government information" means information created, collected, processed, disseminated, or disposed of by or for the Federal Government.

7. "Information dissemination product" means any book, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic, an agency disseminates to the public. This definition includes any electronic document, CD-ROM, or web page.

8. "Dissemination" means agency initiated or sponsored distribution of information to the public (see 5 C.F.R. 1320.3(d) (definition of "Conduct or Sponsor"). Dissemination does not include distribution limited to government employees or agency contractors or grantees; intra- or inter-agency use or sharing of government information; and responses to requests for agency records under the Freedom of Information Act, the Privacy Act, the Federal Advisory Committee Act or other similar law. This definition also does not include distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes.

9. "Influential", when used in the phrase "influential scientific, financial, or statistical information", means that the agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions. Each agency is authorized to define

"influential" in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.

10. "Reproducibility" means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased). If agencies apply the reproducibility test to specific types of original or supporting data, the associated guidelines shall provide relevant definitions of reproducibility (e.g., standards for replication of laboratory data). With respect to analytic results, "capable of being substantially reproduced" means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error.

Return to this article at:
http://www.whitehouse.gov/omb/fedreg/reproducible.html

Click to Print this document

PART 1607——UNIFORM GUIDELINES ON EMPLOYEE SELECTION PROCEDURES (1978)

COMPREHENSIVE TABLE OF CONTENTS

K. Review of validity studies for currency

1607.6. Use of selection procedures which have not been validated

A. Use of alternate selection procedures to eliminate adverse impact

B. Where validity studies cannot or need not be performed

(1) Where informal or unscored procedures are used

(2) Where formal and scored procedures are used

1607.7. Use of other validity studies

A. Validity studies not conducted by the user

B. Use of criterion-related validity evidence from other sources

(1) Validity evidence

(2) Job similarity

(3) Fairness evidence

C. Validity evidence from multiunit study

D. Other significant variables

1607.8. Cooperative studies

A. Encouragement of cooperative studies

B. Standards for use of cooperative studies

1607.9. No assumption of validity

A. Unacceptable substitutes for evidence of validity

B. Encouragement of professional supervision

1607.10. Employment agencies and employment services

A. Where selection procedures are devised by agency

B. Where selection procedures are devised elsewhere

1607.11. Disparate treatment

1607.12. Retesting of applicants

1607.13. Affirmative action

A. Affirmative action obligations

B. Encouragement of voluntary affirmative action programs

TECHNICAL STANDARDS
1607.14. Technical standards for validity studies

A. Validity studies should be based on review of information about the job

B. Technical standards for criterion-related validity studies
- (1) Technical feasibility
- (2) Analysis of the job
- (3) Criterion measures
- (4) Representativeness of the sample
- (5) Statistical relationships
- (6) Operational use of selection procedures
- (7) Overstatement of validity findings
- (8) Fairness
  - (a) Unfairness defined
  - (b) Investigation of fairness
  - (c) General considerations in fairness investigations
  - (d) When unfairness is shown
  - (e) Technical feasibility of fairness studies
  - (f) Continued use of selection procedures when fairness studies not feasible

C. Technical standards for content validity studies
- (1) Appropriateness of content validity studies
- (2) Job analysis for content validity
- (3) Development of selection procedures
- (4) Standards for demonstrating content validity
- (5) Reliability
- (6) Prior training or experience
- (7) Content validity of training success
- (8) Operational use
- (9) Ranking based on content validity studies

D. Technical standards for construct validity studies
- (1) Appropriateness of construct validity studies
- (2) Job analysis for construct validity studies
- (3) Relationship to the job
- (4) Use of construct validity study without new criterion-related evidence

DOCUMENTATION OF IMPACT AND VALIDITY EVIDENCE

1607.15. Documentation of impact and validity evidence

A. Required information
- (1) Simplified recordkeeping for users with less than 100 employees
- (2) Information on impact
  - (a) Collection of information on impact
  - (b) When adverse impact has been eliminated in the total selection process
  - (c) When data insufficient to determine impact
- (3) Documentation of validity evidence
  - (a) Types of evidence
  - (b) Form of report
  - (c) Completeness

B. Criterion-related validity studies
- (1) User(s), location(s), and date(s) of study
- (2) Problem and setting
- (3) Job analysis or review of job information
- (4) Job titles and codes
- (5) Criterion measures
- (6) Sample description
- (7) Description of selection procedures
- (8) Techniques and results
- (9) Alternative procedures investigated
- (10) Uses and applications
- (11) Source data
- (12) Contact person
- (13) Accuracy and completeness

C. Content validity studies
- (1) User(s), location(s), and date(s) of study
- (2) Problem and setting
- (3) Job analysis – Content of the job
- (4) Selection procedure and its content
- (5) Relationship between the selection procedure and the job
- (6) Alternative procedures investigated
- (7) Uses and applications
- (8) Contact person
- (9) Accuracy and completeness

D. Construct validity studies
- (1) User(s), location(s), and date(s) of study

(2) Problem and setting
(3) Construct definition
(4) Job analysis
(5) Job titles and codes
(6) Selection procedure
(7) Relationship to job performance
(8) Alternative procedures investigated
(9) Uses and applications
(10) Accuracy and completeness
(11) Source data
(12) Contact person

E. Evidence of validity from other studies
(1) Evidence from criterion-related validity studies
(a) Job information
(b) Relevance of criteria
(c) Other variables
(d) Use of the selection procedure
(e) Bibliography
(2) Evidence from content validity studies
(3) Evidence from construct validity studies

F. Evidence of validity from cooperative studies G. Selection for higher level job H. Interim use of selection procedures

DEFINITIONS
1607.16. Definitions

APPENDIX
1607.17. Policy statement on affirmative action (see section 13B)
1607.18. Citations

Authority:
Secs. 709 and 713, Civil Rights Act of 1964 (78 Stat. 265) as amended by the Equal Employment Opportunity Act of 1972 (Pub. L. 92—261); 42 U.S.C. 2000e—8, 2000e—12.
Source:
43 FR 38295, 38312, Aug. 25, 1978, unless otherwise noted.
****

General Principles

§§1607.1 Statement of Purpose.

A. Need for uniformity - Issuing agencies.

The Federal government's need for a uniform set of principles on the question of the use of tests and other selection procedures has long been recognized. The Equal Employment Opportunity Commission, the Civil Service Commission, the Department of Labor, and the Department of Justice jointly have adopted these uniform guidelines to meet that need, and to apply the same principles to the Federal Government as are applied to other employers.

B. Purpose of guidelines.

These guidelines incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. They are designed to provide a framework for determining the proper use of tests and other selection procedures. These guidelines do not require a user to conduct validity studies of selection procedures where no adverse impact results. However, all users are encouraged to use selection procedures which are valid, especially users operating under merit principles.

C. Relation to prior guidelines.

These guidelines are based upon and supersede previously issued guidelines on employee selection procedures. These guidelines have been built upon court decisions, the previously issued guidelines of the agencies, and the practical experience of the agencies, as well as the standards of the psychological profession. These guidelines are intended to be consistent with existing law.

§§1607.2 Scope.

A. Application of guidelines.

These guidelines will be applied by the Equal Employment Opportunity Commission in the enforcement of title VII of the Civil Rights Act of 1964, as amended by the Equal Employment Opportunity Act of 1972 (hereinafter ""title VII""); by the Department of Labor, and the contract compliance agencies until the transfer of authority contemplated by the President's Reorganization Plan No. 1 of 1978, in the administration and enforcement of Executive Order 11246, as amended by Executive Order 11375 (hereinafter ""Executive Order 11246""); by the Civil Service Commission and other Federal agencies subject to section 717 of title VII; by the Civil Service Commission in exercising its responsibilities toward State and local governments

under section 208(b)(1) of the Intergovernmental-Personnel Act; by the Department of Justice in exercising its responsibilities under Federal law; by the Office of Revenue Sharing of the Department of the Treasury under the State and Local Fiscal Assistance Act of 1972, as amended; and by any other Federal agency which adopts them.

B. Employment decisions.

These guidelines apply to tests and other selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above.

C. Selection procedures.

These guidelines apply only to selection procedures which are used as a basis for making employment decisions. For example, the use of recruiting procedures designed to attract members of a particular race, sex, or ethnic group, which were previously denied employment opportunities or which are currently underutilized, may be necessary to bring an employer into compliance with Federal law, and is frequently an essential element of any effective affirmative action program; but recruitment practices are not considered by these guidelines to be selection procedures. Similarly, these guidelines do not pertain to the question of the lawfulness of a seniority system within the meaning of section 703(h), Executive Order 11246 or other provisions of Federal law or regulation, except to the extent that such systems utilize selection procedures to determine qualifications or abilities to perform the job. Nothing in these guidelines is intended or should be interpreted as discouraging the use of a selection procedure for the purpose of determining qualifications or for the purpose of selection on the basis of relative qualifications, if the selection procedure had been validated in accord with these guidelines for each such purpose for which it is to be used.

D. Limitations.

These guidelines apply only to persons subject to title VII, Executive Order 11246, or other equal employment opportunity requirements of Federal law. These guidelines do not apply to responsibilities under the Age Discrimination in Employment Act of 1967, as amended, not to discriminate on the basis of age, or under sections 501, 503, and 504 of the Rehabilitation Act of 1973, not to discriminate on the basis of handicap.

E. Indian preference not affected.

These guidelines do not restrict any obligation imposed or right granted by Federal law to users to extend a preference in employment to Indians living on or near an Indian reservation in

connection with employment opportunities on or near an Indian reservation.

§§1607.3 Discrimination defined: Relationship between use of selection procedures and discrimination.

A. Procedure having adverse impact constitutes discrimination unless justified.

The use of any selection procedure which has an adverse impact on the hiring, promotion, or other employment or membership opportunities of members of any race, sex, or ethnic group will be considered to be discriminatory and inconsistent with these guidelines, unless the procedure has been validated in accordance with these guidelines, or the provisions of section 6 below are satisfied.

B. Consideration of suitable alternative selection procedures.

Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. Accordingly, whenever a validity study is called for by these guidelines, the user should include, as a part of the validity study, an investigation of suitable alternative selection procedures and suitable alternative methods of using the selection procedure which have as little adverse impact as possible, to determine the appropriateness of using or validating them in accord with these guidelines. If a user has made a reasonable effort to become aware of such alternative procedures and validity has been demonstrated in accord with these guidelines, the use of the test or other selection procedure may continue until such time as it should reasonably be reviewed for currency. Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines. This subsection is not intended to preclude the combination of procedures into a significantly more valid procedure, if the use of such a combination has been shown to be in compliance with the guidelines.

§§1607.4 Information on impact.

A. Records concerning impact.

Each user should maintain and have available for inspection records or other information which will disclose the impact which its tests and other selection procedures have upon employment opportunities of persons by identifiable race, sex, or ethnic group as set forth in paragraph B of this section, in order to determine compliance with these guidelines. Where there are large numbers of applicants and procedures are administered frequently, such information may be retained on a sample basis, provided that the sample is appropriate in terms of the applicant population and adequate in size.

B. Applicable race, sex, and ethnic groups for recordkeeping.

The records called for by this section are to be maintained by sex, and the following races and ethnic groups: Blacks (Negroes), American Indians (including Alaskan Natives), Asians (including Pacific Islanders), Hispanic (including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), whites (Caucasians) other than Hispanic, and totals. The race, sex, and ethnic classifications called for by this section are consistent with the Equal Employment Opportunity Standard Form 100, Employer Information Report EEO—1 series of reports. The user should adopt safeguards to insure that the records required by this paragraph are used for appropriate purposes such as determining adverse impact, or (where required) for developing and monitoring affirmative action programs, and that such records are not used improperly. See sections 4E and 17(4), below.

C. Evaluation of selection rates.

The ""bottom line."" If the information called for by sections 4A and B above shows that the total selection process for a job has an adverse impact, the individual components of the selection process should be evaluated for adverse impact. If this information shows that the total selection process does not have an adverse impact, the Federal enforcement agencies, in the exercise of their administrative and prosecutorial discretion, in usual circumstances, will not expect a user to evaluate the individual components for adverse impact, or to validate such individual components, and will not take enforcement action based upon adverse impact of any component of that process, including the separate parts of a multipart selection procedure or any separate procedure that is used as an alternative method of selection. However, in the following circumstances the Federal enforcement agencies will expect a user to evaluate the individual components for adverse impact and may, where appropriate, take enforcement action with respect to the individual components: (1) Where the selection procedure is a significant factor in the continuation of patterns of assignments of incumbent employees caused by prior discriminatory employment practices, (2) where the weight of court decisions or administrative interpretations hold that a specific procedure (such as height or weight requirements or no-arrest records) is not job related in the same or similar circumstances. In unusual circumstances, other than those listed in (1) and (2) of this paragraph, the Federal enforcement agencies may request a user to evaluate the individual components for adverse impact and may, where appropriate, take enforcement action with respect to the individual component.

D. Adverse impact and the ""four-fifths rule.""

A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are

significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group. Where the user's evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too small to be reliable, evidence concerning the impact of the procedure over a longer period of time and/or evidence concerning the impact which the selection procedure had when used in the same manner in similar circumstances elsewhere may be considered in determining adverse impact. Where the user has not maintained data on adverse impact as required by the documentation section of applicable guidelines, the Federal enforcement agencies may draw an inference of adverse impact of the selection process from the failure of the user to maintain such data, if the user has an underutilization of a group in the job category, as compared to the group's representation in the relevant labor market or, in the case of jobs filled from within, the applicable work force.

E. Consideration of user's equal employment opportunity posture.

In carrying out their obligations, the Federal enforcement agencies will consider the general posture of the user with respect to equal employment opportunity for the job or group of jobs in question. Where a user has adopted an affirmative action program, the Federal enforcement agencies will consider the provisions of that program, including the goals and timetables which the user has adopted and the progress which the user has made in carrying out that program and in meeting the goals and timetables. While such affirmative action programs may in design and execution be race, color, sex, or ethnic conscious, selection procedures under such programs should be based upon the ability or relative ability to do the work.

(Approved by the Office of Management and Budget under control number 3046—0017)
(Pub. L. 96—511, 94 Stat. 2812 (44 U.S.C. 3501 et seq.))
[43 FR 38295, 38312, Aug. 25, 1978, as amended at 46 FR 63268, Dec. 31, 1981]

§§1607.5 General standards for validity studies.

A. Acceptable types of validity studies.

For the purposes of satisfying these guidelines, users may rely upon criterion-related validity studies, content validity studies or construct validity studies, in accordance with the standards set forth in the technical standards of these guidelines, section 14 below. New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession.

B. Criterion-related, content, and construct validity.

Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. See section 14B below. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. See 14C below. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated. See section 14D below.

C. Guidelines are consistent with professional standards.

The provisions of these guidelines relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (American Psychological Association, Washington, DC, 1974) (hereinafter ""A.P.A. Standards"") and standard textbooks and journals in the field of personnel selection.

D. Need for documentation of validity.

For any selection procedure which is part of a selection process which has an adverse impact and which selection procedure has an adverse impact, each user should maintain and have available such documentation as is described in section 15 below.

E. Accuracy and standardization.

Validity studies should be carried out under conditions which assure insofar as possible the adequacy and accuracy of the research and the report. Selection procedures should be administered and scored under standardized conditions.

F. Caution against selection on basis of knowledges, skills, or ability learned in brief orientation period.

In general, users should avoid making employment decisions on the basis of measures of knowledges, skills, or abilities which are normally learned in a brief orientation period, and which have an adverse impact.

G. Method of use of selection procedures.

The evidence of both the validity and utility of a selection procedure should support the method the user chooses for operational use of the procedure, if that method of use has a greater adverse impact than another method of use. Evidence which may be sufficient to support the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the same procedure on a ranking basis under these guidelines. Thus, if a user decides to use a selection procedure on a ranking basis, and that method of use has a greater adverse impact than use on an appropriate pass/fail basis (see section 5H below), the user should have sufficient evidence of validity and utility to support the use on a ranking basis. See sections 3B, 14B (5) and (6), and 14C (8) and (9).

H. Cutoff scores.

Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff score than appropriate in light of such expectations have little or no chance of being selected for employment, the higher cutoff score may be appropriate, but the degree of adverse impact should be considered.

I. Use of selection procedures for higher level jobs.

If job progression structures are so established that employees will probably, within a reasonable period of time and in a majority of cases, progress to a higher level, it may be considered that the applicants are being evaluated for a job or jobs at the higher level. However, where job progression is not so nearly automatic, or the time span is such that higher level jobs or employees'' potential may be expected to change in significant ways, it should be considered that applicants are being evaluated for a job at or near the entry level. A ""reasonable period of time"" will vary for different jobs and employment situations but will seldom be more than 5 years. Use of selection procedures to evaluate applicants for a higher level job would not be appropriate: (1) If the majority of those remaining employed do not progress to the higher level job; (2) If there is a reason to doubt that the higher level job will continue to require essentially similar skills during the progression period; or (3) If the selection procedures measure knowledges, skills, or abilities required for advancement which would be expected to develop principally from the training or experience on the job.

J. Interim use of selection procedures.

Users may continue the use of a selection procedure which is not at the moment fully supported by the required evidence of validity, provided: (1) The user has available substantial evidence of validity, and (2) the user has in progress, when technically feasible, a study which is designed to produce the additional evidence required by these guidelines within a reasonable time. If such a

study is not technically feasible, see section 6B. If the study does not demonstrate validity, this provision of these guidelines for interim use shall not constitute a defense in any action, nor shall it relieve the user of any obligations arising under Federal law.

K. Review of validity studies for currency.

Whenever validity has been shown in accord with these guidelines for the use of a particular selection procedure for a job or group of jobs, additional studies need not be performed until such time as the validity study is subject to review as provided in section 3B above. There are no absolutes in the area of determining the currency of a validity study. All circumstances concerning the study, including the validation strategy used, and changes in the relevant labor market and the job should be considered in the determination of when a validity study is outdated.

§§1607.6 Use of selection procedures which have not been validated.

A. Use of alternate selection procedures to eliminate adverse impact.

A user may choose to utilize alternative selection procedures in order to eliminate adverse impact or as part of an affirmative action program. See section 13 below. Such alternative procedures should eliminate the adverse impact in the total selection process, should be lawful and should be as job related as possible.

B. Where validity studies cannot or need not be performed.

There are circumstances in which a user cannot or need not utilize the validation techniques contemplated by these guidelines. In such circumstances, the user should utilize selection procedures which are as job related as possible and which will minimize or eliminate adverse impact, as set forth below.

(1) Where informal or unscored procedures are used.

When an informal or unscored selection procedure which has an adverse impact is utilized, the user should eliminate the adverse impact, or modify the procedure to one which is a formal, scored or quantified measure or combination of measures and then validate the procedure in accord with these guidelines, or otherwise justify continued use of the procedure in accord with Federal law.

(2) Where formal and scored procedures are used.

When a formal and scored selection procedure is used which has an adverse impact, the validation techniques contemplated by these guidelines usually should be followed if technically feasible. Where the user cannot or need not follow the validation techniques anticipated by these

guidelines, the user should either modify the procedure to eliminate adverse impact or otherwise justify continued use of the procedure in accord with Federal law.

§§1607.7 Use of other validity studies.

A. Validity studies not conducted by the user.

Users may, under certain circumstances, support the use of selection procedures by validity studies conducted by other users or conducted by test publishers or distributors and described in test manuals. While publishers of selection procedures have a professional obligation to provide evidence of validity which meets generally accepted professional standards (see section 5C above), users are cautioned that they are responsible for compliance with these guidelines. Accordingly, users seeking to obtain selection procedures from publishers and distributors should be careful to determine that, in the event the user becomes subject to the validity requirements of these guidelines, the necessary information to support validity has been determined and will be made available to the user.

B. Use of criterion-related validity evidence from other sources.

Criterion-related validity studies conducted by one test user, or described in test manuals and the professional literature, will be considered acceptable for use by another user when the following requirements are met:

(1) Validity evidence.

Evidence from the available studies meeting the standards of section 14B below clearly demonstrates that the selection procedure is valid;

(2) Job similarity.

The incumbents in the user's job and the incumbents in the job or group of jobs on which the validity study was conducted perform substantially the same major work behaviors, as shown by appropriate job analyses both on the job or group of jobs on which the validity study was performed and on the job for which the selection procedure is to be used; and

(3) Fairness evidence.

The studies include a study of test fairness for each race, sex, and ethnic group which constitutes a significant factor in the borrowing user's relevant labor market for the job or jobs in question. If the studies under consideration satisfy paragraphs (1) and (2) of this paragraph B., 1/4 above but do not contain an investigation of test fairness, and it is not technically feasible for the borrowing user to conduct an internal study of test fairness, the borrowing user may utilize the study until studies conducted elsewhere meeting the requirements of these guidelines show

test unfairness, or until such time as it becomes technically feasible to conduct an internal study of test fairness and the results of that study can be acted upon. Users obtaining selection procedures from publishers should consider, as one factor in the decision to purchase a particular selection procedure, the availability of evidence concerning test fairness.

C. Validity evidence from multi-unit study.

If validity evidence from a study covering more than one unit within an organization satisfies the requirements of section 14B below, evidence of validity specific to each unit will not be required unless there are variables which are likely to affect validity significantly.

D. Other significant variables.

If there are variables in the other studies which are likely to affect validity significantly, the user may not rely upon such studies, but will be expected either to conduct an internal validity study or to comply with section 6 above.

§§1607.8 Cooperative studies.

A. Encouragement of cooperative studies.

The agencies issuing these guidelines encourage employers, labor organizations, and employment agencies to cooperate in research, development, search for lawful alternatives, and validity studies in order to achieve procedures which are consistent with these guidelines.

B. Standards for use of cooperative studies.

If validity evidence from a cooperative study satisfies the requirements of section 14 below, evidence of validity specific to each user will not be required unless there are variables in the user's situation which are likely to affect validity significantly.

§§1607.9 No assumption of validity.

A. Unacceptable substitutes for evidence of validity.

Under no circumstances will the general reputation of a test or other selection procedures, its author or its publisher, or casual reports of it's validity be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on a procedure's name or descriptive labels; all forms of promotional literature; data bearing on the frequency of a procedure's usage; testimonial statements and credentials of sellers, users, or consultants; and other non-empirical or anecdotal accounts of selection practices or selection outcomes.

B. Encouragement of professional supervision.

Professional supervision of selection activities is encouraged but is not a substitute for documented evidence of validity. The enforcement agencies will take into account the fact that a thorough job analysis was conducted and that careful development and use of a selection procedure in accordance with professional standards enhance the probability that the selection procedure is valid for the job.

§§1607.10 Employment agencies and employment services.

A. Where selection procedures are devised by agency.

An employment agency, including private employment agencies and State employment agencies, which agrees to a request by an employer or labor organization to device and utilize a selection procedure should follow the standards in these guidelines for determining adverse impact. If adverse impact exists the agency should comply with these guidelines. An employment agency is not relieved of its obligation herein because the user did not request such validation or has requested the use of some lesser standard of validation than is provided in these guidelines. The use of an employment agency does not relieve an employer or labor organization or other user of its responsibilities under Federal law to provide equal employment opportunity or its obligations as a user under these guidelines.

B. Where selection procedures are devised elsewhere.

Where an employment agency or service is requested to administer a selection procedure which has been devised elsewhere and to make referrals pursuant to the results, the employment agency or service should maintain and have available evidence of the impact of the selection and referral procedures which it administers. If adverse impact results the agency or service should comply with these guidelines. If the agency or service seeks to comply with these guidelines by reliance upon validity studies or other data in the possession of the employer, it should obtain and have available such information.

§§1607.11 Disparate treatment.

The principles of disparate or unequal treatment must be distinguished from the concepts of validation. A selection procedure – even though validated against job performance in accordance with these guideline – cannot be imposed upon members of a race, sex, or ethnic group where other employees, applicants, or members have not been subjected to that standard. Disparate treatment occurs where members of a race, sex, or ethnic group have been denied the same employment, promotion, membership, or other employment opportunities as have been available to other employees or applicants. Those employees or applicants who have been denied equal treatment, because of prior discriminatory practices or policies, must at least be afforded the same opportunities as had existed for other employees or applicants during the period of discrimination. Thus, the persons who were in the class of persons discriminated against during the period the user followed the discriminatory practices should be allowed the opportunity to

qualify under less stringent selection procedures previously followed, unless the user demonstrates that the increased standards are required by business necessity. This section does not prohibit a user who has not previously followed merit standards from adopting merit standards which are in compliance with these guidelines; nor does it preclude a user who has previously used invalid or unvalidated selection procedures from developing and using procedures which are in accord with these guidelines.

§§1607.12 Retesting of applicants.

Users should provide a reasonable opportunity for retesting and reconsideration. Where examinations are administered periodically with public notice, such reasonable opportunity exists, unless persons who have previously been tested are precluded from retesting. The user may however take reasonable steps to preserve the security of its procedures.

§§1607.13 Affirmative action.

A. Affirmative action obligations.

The use of selection procedures which have been validated pursuant to these guidelines does not relieve users of any obligations they may have to undertake affirmative action to assure equal employment opportunity. Nothing in these guidelines is intended to preclude the use of lawful selection procedures which assist in remedying the effects of prior discriminatory practices, or the achievement of affirmative action objectives.

B. Encouragement of voluntary affirmative action programs.

These guidelines are also intended to encourage the adoption and implementation of voluntary affirmative action programs by users who have no obligation under Federal law to adopt them; but are not intended to impose any new obligations in that regard. The agencies issuing and endorsing these guidelines endorse for all private employers and reaffirm for all governmental employers the Equal Employment Opportunity Coordinating Council's ""Policy Statement on Affirmative Action Programs for State and Local Government Agencies"" (41 FR 38814, September 13, 1976). That policy statement is attached hereto as appendix, section 17. Technical Standards

§§1607.14 Technical standards for validity studies.

The following minimum standards, as applicable, should be met in conducting a validity study. Nothing in these guidelines is intended to preclude the development and use of other professionally acceptable techniques with respect to validation of selection procedures. Where it is not technically feasible for a user to conduct a validity study, the user has the obligation otherwise to comply with these guidelines. See sections 6 and 7 above.

A. Validity studies should be based on review of information about the job.

Any validity study should be based upon a review of information about the job for which the selection procedure is to be used. The review should include a job analysis except as provided in section 14B(3) below with respect to criterion-related validity. Any method of job analysis may be used if it provides the information required for the specific validation strategy used.

B. Technical standards for criterion-related validity studies.

(1) Technical feasibility.

Users choosing to validate a selection procedure by a criterion-related validity strategy should determine whether it is technically feasible (as defined in section 16) to conduct such a study in the particular employment context. The determination of the number of persons necessary to permit the conduct of a meaningful criterion-related study should be made by the user on the basis of all relevant information concerning the selection procedure, the potential sample and the employment situation. Where appropriate, jobs with substantially the same major work behaviors may be grouped together for validity studies, in order to obtain an adequate sample. These guidelines do not require a user to hire or promote persons for the purpose of making it possible to conduct a criterion-related study.

(2) Analysis of the job.

There should be a review of job information to determine measures of work behavior(s) or performance that are relevant to the job or group of jobs in question. These measures or criteria are relevant to the extent that they represent critical or important job duties, work behaviors or work outcomes as developed from the review of job information. The possibility of bias should be considered both in selection of the criterion measures and their application. In view of the possibility of bias in subjective evaluations, supervisory rating techniques and instructions to raters should be carefully developed. All criterion measures and the methods for gathering data need to be examined for freedom from factors which would unfairly alter scores of members of any group. The relevance of criteria and their freedom from bias are of particular concern when there are significant differences in measures of job performance for different groups.

(3) Criterion measures.

Proper safeguards should be taken to insure that scores on selection procedures do not enter into any judgments of employee adequacy that are to be used as criterion measures. Whatever criteria are used should represent important or critical work behavior(s) or work outcomes. Certain criteria may be used without a full job analysis if the user can show the importance of the criteria to the particular employment context. These criteria include but are not limited to production rate, error rate, tardiness, absenteeism, and length of service. A

standardized rating of overall work performance may be used where a study of the job shows that it is an appropriate criterion. Where performance in training is used as a criterion, success in training should be properly measured and the relevance of the training should be shown either through a comparsion of the content of the training program with the critical or important work behavior(s) of the job(s), or through a demonstration of the relationship between measures of performance in training and measures of job performance. Measures of relative success in training include but are not limited to instructor evaluations, performance samples, or tests. Criterion measures consisting of paper and pencil tests will be closely reviewed for job relevance.

(4) Representativeness of the sample.

Whether the study is predictive or concurrent, the sample subjects should insofar as feasible be representative of the candidates normally available in the relevant labor market for the job or group of jobs in question, and should insofar as feasible include the races, sexes, and ethnic groups normally available in the relevant job market. In determining the representativeness of the sample in a concurrent validity study, the user should take into account the extent to which the specific knowledges or skills which are the primary focus of the test are those which employees learn on the job. Where samples are combined or compared, attention should be given to see that such samples are comparable in terms of the actual job they perform, the length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or that these factors are included in the design of the study and their effects identified.

(5) Statistical relationships.

The degree of relationship between selection procedure scores and criterion measures should be examined and computed, using professionally acceptable statistical procedures. Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection procedure and job performance should not necessarily discourage other investigations of the validity of that selection procedure.

(6) Operational use of selection procedures.

Users should evaluate each selection procedure to assure that it is appropriate for operational use, including establishment of cutoff scores or rank ordering. Generally, if other factors remain the same, the greater the magnitude of the relationship (e.g., correlation coefficient) between performance on a selection procedure and one or more criteria of performance on the job, and the greater the importance and number of aspects of job performance

covered by the criteria, the more likely it is that the procedure will be appropriate for use. Reliance upon a selection procedure which is significantly related to a criterion measure, but which is based upon a study involving a large number of subjects and has a low correlation coefficient will be subject to close review if it has a large adverse impact. Sole reliance upon a single selection instrument which is related to only one of many job duties or aspects of job performance will also be subject to close review. The appropriateness of a selection procedure is best evaluated in each particular situation and there are no minimum correlation coefficients applicable to all employment situations. In determining whether a selection procedure is appropriate for operational use the following considerations should also be taken into account: The degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity.

(7) Overstatement of validity findings.

Users should avoid reliance upon techniques which tend to overestimate validity findings as a result of capitalization on chance unless an appropriate safeguard is taken. Reliance upon a few selection procedures or criteria of successful job performance when many selection procedures or criteria of performance have been studied, or the use of optimal statistical weights for selection procedures computed in one sample, are techniques which tend to inflate validity estimates as a result of chance. Use of a large sample is one safeguard: cross-validation is another.

(8) Fairness.

This section generally calls for studies of unfairness where technically feasible. The concept of fairness or unfairness of selection procedures is a developing concept. In addition, fairness studies generally require substantial numbers of employees in the job or group of jobs being studied. For these reasons, the Federal enforcement agencies recognize that the obligation to conduct studies of fairness imposed by the guidelines generally will be upon users or groups of users with a large number of persons in a a job class, or test developers; and that small users utilizing their own selection procedures will generally not be obligated to conduct such studies because it will be technically infeasible for them to do so.

(a) Unfairness defined. When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.

(b) Investigation of fairness. Where a selection procedure results in an adverse impact on a race, sex, or ethnic group identified in accordance with the classifications set forth in section 4 above and that group is a significant factor in the relevant labor market, the user generally should investigate the possible existence of unfairness for that group if it is

technically feasible to do so. The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness. Where the weight of evidence from other studies shows that the selection procedure predicts fairly for the group in question and for the same or similar jobs, such evidence may be relied on in connection with the selection procedure at issue.

(c) General considerations in fairness investigations. Users conducting a study of fairness should review the A.P.A. Standards regarding investigation of possible bias in testing. An investigation of fairness of a selection procedure depends on both evidence of validity and the manner in which the selection procedure is to be used in a particular employment context. Fairness of a selection procedure cannot necessarily be specified in advance without investigating these factors. Investigation of fairness of a selection procedure in samples where the range of scores on selection procedures or criterion measures is severely restricted for any subgroup sample (as compared to other subgroup samples) may produce misleading evidence of unfairness. That factor should accordingly be taken into account in conducting such studies and before reliance is placed on the results.

(d) When unfairness is shown. If unfairness is demonstrated through a showing that members of a particular group perform better or poorer on the job than their scores on the selection procedure would indicate through comparison with how members of other groups perform, the user may either revise or replace the selection instrument in accordance with these guidelines, or may continue to use the selection instrument operationally with appropriate revisions in its use to assure compatibility between the probability of successful job performance and the probability of being selected.

(e) Technical feasibility of fairness studies. In addition to the general conditions needed for technical feasibility for the conduct of a criterion-related study (see section 16, below) an investigation of fairness requires the following:

> **(i)** An adequate sample of persons in each group available for the study to achieve findings of statistical significance. Guidelines do not require a user to hire or promote persons on the basis of group classifications for the purpose of making it possible to conduct a study of fairness; but the user has the obligation otherwise to comply with these guidelines.
>
> **(ii)** The samples for each group should be comparable in terms of the actual job they perform, length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or such factors should be included in the design of the study and their effects identified.

(f) Continued use of selection procedures when fairness studies not feasible. If a study of fairness should otherwise be performed, but is not technically feasible, a selection procedure may be used which has otherwise met the validity standards of these guidelines, unless the technical infeasibility resulted from discriminatory employment practices which are demonstrated by facts other than past failure to conform with requirements for validation of selection procedures. However, when it becomes technically feasible for the user to perform a study of fairness and such a study is otherwise called for, the user should conduct the study of fairness.

C. Technical standards for content validity studies.

(1) Appropriateness of content validity studies.

Users choosing to validate a selection procedure by a content validity strategy should determine whether it is appropriate to conduct such a study in the particular employment context. A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined as provided in section 14C(4) below, and if that knowledge, skill, or ability is a necessary prerequisite to successful job performance.

A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability. Content validity is also not an appropriate strategy when the selection procedure involves knowledges, skills, or abilities which an employee will be expected to learn on the job.

(2) Job analysis for content validity.

There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the job analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

(3) Development of selection procedures.

A selection procedure designed to measure the work behavior may be developed specifically from the job and job analysis in question, or may have been previously developed by

the user, or by other users or by a test publisher.

(4) Standards for demonstrating content validity.

To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job. In the case of a selection procedure measuring a knowledge, skill, or ability, the knowledge, skill, or ability being measured should be operationally defined. In the case of a selection procedure measuring a knowledge, the knowledge being measured should be operationally defined as that body of learned information which is used in and is a necessary prerequisite for observable aspects of work behavior of the job. In the case of skills or abilities, the skill or ability being measured should be operationally defined in terms of observable aspects of work behavior of the job. For any selection procedure measuring a knowledge, skill, or ability the user should show that (a) the selection procedure measures and is a representative sample of that knowledge, skill, or ability; and (b) that knowledge, skill, or ability is used in and is a necessary prerequisite to performance of critical or important work behavior(s).

In addition, to be content valid, a selection procedure measuring a skill or ability should either closely approximate an observable work behavior, or its product should closely approximate an observable work product. If a test purports to sample a work behavior or to provide a sample of a work product, the manner and setting of the selection procedure and its level and complexity should closely approximate the work situation. The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity. As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.

(5) Reliability.

The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure.

(6) Prior training or experience.

A requirement for or evaluation of specific prior training or experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the content of the job for which the training or experience is to be required or evaluated. The critical consideration is the resemblance between the specific behaviors, products, knowledges, skills, or abilities in the experience or training and the specific behaviors, products, knowledges, skills, or

abilities required on the job, whether or not there is close resemblance between the experience or training as a whole and the job as a whole.

(7) Content validity of training success.

Where a measure of success in a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, the use should be justified on the relationship between the content of the training program and the content of the job.

(8) Operational use.

A selection procedure which is supported on the basis of content validity may be used for a job if it represents a critical work behavior (i.e., a behavior which is necessary for performance of the job) or work behaviors which constitute most of the important parts of the job.

(9) Ranking based on content validity studies.

If a user can show, by a job analysis or otherwise, that a higher score on a content valid selection procedure is likely to result in better job performance, the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance.

D. Technical standards for construct validity studies.

(1) Appropriateness of construct validity studies.

Construct validity is a more complex strategy than either criterion-related or content validity. Construct validation is a relatively new and developing procedure in the employment field, and there is at present a lack of substantial literature extending the concept to employment practices. The user should be aware that the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which include criterion related validity studies and which may include content validity studies. Users choosing to justify use of a selection procedure by this strategy should therefore take particular care to assure that the validity study meets the standards set forth below.

(2) Job analysis for construct validity studies.

There should be a job analysis. This job analysis should show the work behavior(s) required for successful performance of the job, or the groups of jobs being studied, the critical or important work behavior(s) in the job or group of jobs being studied, and an identification of the construct(s) believed to underlie successful performance of these critical or important work

behaviors in the job or jobs in question. Each construct should be named and defined, so as to distinguish it from other constructs. If a group of jobs is being studied the jobs should have in common one or more critical or important work behav- iors at a comparable level of complexity.

(3) Relationship to the job.

A selection procedure should then be identified or developed which measures the construct identified in accord with subparagraph (2) above. The user should show by empirical evidence that the selection procedure is validly related to the construct and that the construct is validly related to the performance of critical or important work behavior(s). The relationship between the construct as measured by the selection procedure and the related work behavior(s) should be supported by empirical evidence from one or more criterion-related studies involving the job or jobs in question which satisfy the provisions of section 14B above.

(4) Use of construct validity study without new criterion-related evidence.

(a) Standards for use. Until such time as professional literature provides more guidance on the use of construct validity in employment situations, the Federal agencies will accept a claim of construct validity without a criterion-related study which satisfies section 14B above only when the selection procedure has been used elsewhere in a situation in which a criterion-related study has been conducted and the use of a criterion-related validity study in this context meets the standards for transportability of criterion-related validity studies as set forth above in section 7. However, if a study pertains to a number of jobs having common critical or important work behaviors at a comparable level of complexity, and the evidence satisfies subparagraphs 14B (2) and (3) above for those jobs with criterion-related validity evidence for those jobs, the selection procedure may be used for all the jobs to which the study pertains. If construct validity is to be generalized to other jobs or groups of jobs not in the group studied, the Federal enforcement agencies will expect at a minimum additional empirical research evidence meeting the standards of subparagraphs section 14B (2) and (3) above for the additional jobs or groups of jobs.

(b) Determination of common work behaviors. In determining whether two or more jobs have one or more work behavior(s) in common, the user should compare the observed work behavior(s) in each of the jobs and should compare the observed work product(s) in each of the jobs. If neither the observed work behavior(s) in each of the jobs nor the observed work product(s) in each of the jobs are the same, the Federal enforcement agencies will presume that the work behavior(s) in each job are different. If the work behaviors are not observable, then evidence of similarity of work products and any other relevant research evidence will be considered in determining whether the work behavior(s) in the two jobs are the same.

§§1607.15 Documentation of impact and validity evidence.

A. Required information.

Users of selection procedures other than those users complying with section 15A(1) below should maintain and have available for each job information on adverse impact of the selection process for that job and, where it is determined a selection process has an adverse impact, evidence of validity as set forth below.

(1) Simplified recordkeeping for users with less than 100 employees.

In order to minimize recordkeeping burdens on employers who employ one hundred (100) or fewer employees, and other users not required to file EEO—1, et seq., reports, such users may satisfy the requirements of this section 15 if they maintain and have available records showing, for each year:

(a) The number of persons hired, promoted, and terminated for each job, by sex, and where appropriate by race and national origin;

(b) The number of applicants for hire and promotion by sex and where appropriate by race and national origin; and

(c) The selection procedures utilized (either standardized or not standardized). These records should be maintained for each race or national origin group (see section 4 above) constituting more than two percent (2%) of the labor force in the relevant labor area. However, it is not necessary to maintain records by race and/or national origin (see §§4 above) if one race or national origin group in the relevant labor area constitutes more than ninety-eight percent (98%) of the labor force in the area. If the user has reason to believe that a selection procedure has an adverse impact, the user should maintain any available evidence of validity for that procedure (see sections 7A and 8).

(2) Information on impact.

(a) Collection of information on impact. Users of selection procedures other than those complying with section 15A(1) above should maintain and have available for each job records or other information showing whether the total selection process for that job has an adverse impact on any of the groups for which records are called for by sections 4B above. Adverse impact determinations should be made at least annually for each such group which constitutes at least 2 percent of the labor force in the relevant labor area or 2 percent of the applicable workforce. Where a total selection process for a job has an adverse impact, the user should maintain and have available records or other information showing which components have an adverse impact. Where the total selection process for a job does not have an adverse impact, information need not be maintained for individual

components except in circumstances set forth in subsection 15A(2)(b) below. If the determination of adverse impact is made using a procedure other than the ""four-fifths rule,"" as defined in the first sentence of section 4D above, a justification, consistent with section 4D above, for the procedure used to determine adverse impact should be available.

(b) When adverse impact has been eliminated in the total selection process. Whenever the total selection process for a particular job has had an adverse impact, as defined in section 4 above, in any year, but no longer has an adverse impact, the user should maintain and have available the information on individual components of the selection process required in the preceding paragraph for the period in which there was adverse impact. In addition, the user should continue to collect such information for at least two (2) years after the adverse impact has been eliminated.

(c) When data insufficient to determine impact. Where there has been an insufficient number of selections to determine whether there is an adverse impact of the total selection process for a particular job, the user should continue to collect, maintain and have available the information on individual components of the selection process required in section 15(A)(2)(a) above until the information is sufficient to determine that the overall selection process does not have an adverse impact as defined in section 4 above, or until the job has changed substantially.

(3) Documentation of validity evidence.

(a) Types of evidence. Where a total selection process has an adverse impact (see section 4 above) the user should maintain and have available for each component of that process which has an adverse impact, one or more of the following types of documentation evidence:

(i) Documentation evidence showing criterion-related validity of the selection procedure (see section 15B, below).

(ii) Documentation evidence showing content validity of the selection procedure (see section 15C, below).

(iii) Documentation evidence showing construct validity of the selection procedure (see section 15D, below).

(iv) Documentation evidence from other studies showing validity of the selection procedure in the user's facility (see section 15E, below).

(v) Documentation evidence showing why a validity study cannot or need not be performed and why continued use of the procedure is consistent with Federal law.

(b) Form of report. This evidence should be compiled in a reasonably complete and organized manner to permit direct evaluation of the validity of the selection procedure. Previously written employer or consultant reports of validity, or reports describing validity studies completed before the issuance of these guidelines are acceptable if they are complete in regard to the documentation requirements contained in this section, or if they satisfied requirements of guidelines which were in effect when the validity study was completed. If they are not complete, the required additional documentation should be appended. If necessary information is not available the report of the validity study may still be used as documentation, but its adequacy will be evaluated in terms of compliance with the requirements of these guidelines.

(c) Completeness. In the event that evidence of validity is reviewed by an enforcement agency, the validation reports completed after the effective date of these guidelines are expected to contain the information set forth below. Evidence denoted by use of the word ""(Essential)"" is considered critical. If information denoted essential is not included, the report will be considered incomplete unless the user affirmatively demonstrates either its unavailability due to circumstances beyond the user's control or special circumstances of the user's study which make the information irrelevant. Evidence not so denoted is desirable but its absence will not be a basis for considering a report incomplete. The user should maintain and have available the information called for under the heading ""Source Data"" in sections 15B(11) and 15D(11). While it is a necessary part of the study, it need not be submitted with the report. All statistical results should be organized and presented in tabular or graphic form to the extent feasible.

B. Criterion-related validity studies.

Reports of criterion-related validity for a selection procedure should include the following information:

(1) User(s), location(s), and date(s) of study.

Dates and location(s) of the job analysis or review of job information, the date(s) and location(s) of the administration of the selection procedures and collection of criterion data, and the time between collection of data on selection procedures and criterion measures should be provided (Essential). If the study was conducted at several locations, the address of each location, including city and State, should be shown.

(2) Problem and setting.

An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing selection procedures and cutoff scores, if any, should be provided.

(3) Job analysis or review of job information.

A description of the procedure used to analyze the job or group of jobs, or to review the job information should be provided (Essential). Where a review of job information results in criteria which may be used without a full job analysis (see section 14B(3)), the basis for the selection of these criteria should be reported (Essential). Where a job analysis is required a complete description of the work behavior(s) or work outcome(s), and measures of their criticality or importance should be provided (Essential). The report should describe the basis on which the behavior(s) or outcome(s) were determined to be critical or important, such as the proportion of time spent on the respective behaviors, their level of difficulty, their frequency of performance, the consequences of error, or other appropriate factors (Essential). Where two or more jobs are grouped for a validity study, the information called for in this subsection should be provided for each of the jobs, and the justification for the grouping (see section 14B(1)) should be provided (Essential).

(4) Job titles and codes.

It is desirable to provide the user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from U.S. Employment Service's Dictionary of Occupational Titles.

(5) Criterion measures.

The bases for the selection of the criterion measures should be provided, together with references to the evidence considered in making the selection of criterion measures (essential). A full description of all criteria on which data were collected and means by which they were observed, recorded, evaluated, and quantified, should be provided (essential). If rating techniques are used as criterion measures, the appraisal form(s) and instructions to the rater(s) should be included as part of the validation evidence, or should be explicitly described and available (essential). All steps taken to insure that criterion measures are free from factors which would unfairly alter the scores of members of any group should be described (essential).

(6) Sample description.

A description of how the research sample was identified and selected should be included (essential). The race, sex, and ethnic composition of the sample, including those groups set forth in section 4A above, should be described (essential). This description should include the size of each subgroup (essential). A description of how the research sample compares with the relevant labor market or work force, the method by which the relevant labor market or work force was defined, and a discussion of the likely effects on validity of differences between the sample and the relevant labor market or work force, are also desirable. Descriptions of educational levels, length of service, and age are also desirable.

(7) Description of selection procedures.

Any measure, combination of measures, or procedure studied should be completely and explicitly described or attached (essential). If commercially available selection procedures are studied, they should be described by title, form, and publisher (essential). Reports of reliability estimates and how they were established are desirable.

(8) Techniques and results.

Methods used in analyzing data should be described (essential). Measures of central tendency (e.g., means) and measures of dispersion (e.g., standard deviations and ranges) for all selection procedures and all criteria should be reported for each race, sex, and ethnic group which constitutes a significant factor in the relevant labor market (essential). The magnitude and direction of all relationships between selection procedures and criterion measures investigated should be reported for each relevant race, sex, and ethnic group and for the total group (essential). Where groups are too small to obtain reliable evidence of the magnitude of the relationship, need not be reported separately. Statements regarding the statistical significance of results should be made (essential). Any statistical adjustments, such as for less then perfect reliability or for restriction of score range in the selection procedure or criterion should be described and explained; and uncorrected correlation coefficients should also be shown (essential). Where the statistical technique categorizes continuous data, such as biserial correlation and the phi coefficient, the categories and the bases on which they were determined should be described and explained (essential). Studies of test fairness should be included where called for by the requirements of section 14B(8) (essential). These studies should include the rationale by which a selection procedure was determined to be fair to the group(s) in question. Where test fairness or unfairness has been demonstrated on the basis of other studies, a bibliography of the relevant studies should be included (essential). If the bibliography includes unpublished studies, copies of these studies, or adequate abstracts or summaries, should be attached (essential). Where revisions have been made in a selection procedure to assure compatability between successful job performance and the probability of being selected, the studies underlying such revisions should be included (essential). All statistical results should be organized and presented by relevant race, sex, and ethnic group (essential).

(9) Alternative procedures investigated.

The selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of the investigation, and the conclusions reached in light of the findings, should be fully described (essential).

(10) Uses and applications.

The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other procedures in a battery) and

available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use, and the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If weights are assigned to different parts of the selection procedure, these weights and the validity of the weighted composite should be reported (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential).

(11) Source data.

Each user should maintain records showing all pertinent information about individual sample members and raters where they are used, in studies involving the validation of selection procedures. These records should be made available upon request of a compliance agency. In the case of individual sample members these data should include scores on the selection procedure(s), scores on criterion measures, age, sex, race, or ethnic group status, and experience on the specific job on which the validation study was conducted, and may also include such things as education, training, and prior job experience, but should not include names and social security numbers. Records should be maintained which show the ratings given to each sample member by each rater.

(12) Contact person.

The name, mailing address, and telephone number of the person who may be contacted for further information about the validity study should be provided (essential).

(13) Accuracy and completeness.

The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

C. Content validity studies.

Reports of content validity for a selection procedure should include the following information:

(1) User(s), location(s) and date(s) of study.

Dates and location(s) of the job analysis should be shown (essential).

(2) Problem and setting.

An explicit definition of the purpose(s) of the study and the circumstances in which the

study was conducted should be provided. A description of existing selection procedures and cutoff scores, if any, should be provided.

(3) Job analysis.

Content of the job. A description of the method used to analyze the job should be provided (essential). The work behavior(s), the associated tasks, and, if the behavior results in a work product, the work products should be completely described (essential). Measures of criticality and/or importance of the work behavior(s) and the method of determining these measures should be provided (essential). Where the job analysis also identified the knowledges, skills, and abilities used in work behavior(s), an operational definition for each knowledge in terms of a body of learned information and for each skill and ability in terms of observable behaviors and outcomes, and the relationship between each knowledge, skill, or ability and each work behavior, as well as the method used to determine this relationship, should be provided (essential). The work situation should be described, including the setting in which work behavior(s) are performed, and where appropriate, the manner in which knowledges, skills, or abilities are used, and the complexity and difficulty of the knowledge, skill, or ability as used in the work behavior(s).

(4) Selection procedure and its content.

Selection procedures, including those constructed by or for the user, specific training requirements, composites of selection procedures, and any other procedure supported by content validity, should be completely and explicitly described or attached (essential). If commercially available selection procedures are used, they should be described by title, form, and publisher (essential). The behaviors measured or sampled by the selection procedure should be explicitly described (essential). Where the selection procedure purports to measure a knowledge, skill, or ability, evidence that the selection procedure measures and is a representative sample of the knowledge, skill, or ability should be provided (essential).

(5) Relationship between the selection procedure and the job.

The evidence demonstrating that the selection procedure is a representative work sample, a representative sample of the work behavior(s), or a representative sample of a knowledge, skill, or ability as used as a part of a work behavior and necessary for that behavior should be provided (essential). The user should identify the work behavior(s) which each item or part of the selection procedure is intended to sample or measure (essential). Where the selection procedure purports to sample a work behavior or to provide a sample of a work product, a comparison should be provided of the manner, setting, and the level of complexity of the selection procedure with those of the work situation (essential). If any steps were taken to reduce adverse impact on a race, sex, or ethnic group in the content of the procedure or in its administration, these steps should be described. Establishment of time limits, if any, and how these limits are related to the speed with which duties must be performed on the job, should be explained. Measures of central tend- ency

(e.g., means) and measures of dispersion (e.g., standard deviations) and estimates of reliability should be reported for all selection procedures if available. Such reports should be made for relevant race, sex, and ethnic subgroups, at least on a statistically reliable sample basis.

(6) Alternative procedures investigated.

The alternative selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of the investigation, and the conclusions reached in light of the findings, should be fully described (essential).

(7) Uses and applications.

The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other procedures in a battery) and available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use, and the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential). In addition, if the selection procedure is to be used for ranking, the user should specify the evidence showing that a higher score on the selection procedure is likely to result in better job performance.

(8) Contact person.

The name, mailing address, and telephone number of the person who may be contacted for further information about the validity study should be provided (essential).

(9) Accuracy and completeness.

The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

D. Construct validity studies.

Reports of construct validity for a selection procedure should include the following information:

(1) User(s), location(s), and date(s) of study.

Date(s) and location(s) of the job analysis and the gathering of other evidence called for by these guidelines should be provided (essential).

(2) Problem and setting.

An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing selection procedures and cutoff scores, if any, should be provided.

(3) Construct definition.

A clear definition of the construct(s) which are believed to underlie successful performance of the critical or important work behavior(s) should be provided (essential). This definition should include the levels of construct performance relevant to the job(s) for which the selection procedure is to be used (essential). There should be a summary of the position of the construct in the psychological literature, or in the absence of such a position, a description of the way in which the definition and measurement of the construct was developed and the psychological theory underlying it (essential). Any quantitative data which identify or define the job constructs, such as factor analyses, should be provided (essential).

(4) Job analysis.

A description of the method used to analyze the job should be provided (essential). A complete description of the work behavior(s) and, to the extent appropriate, work outcomes and measures of their criticality and/or importance should be provided (essential). The report should also describe the basis on which the behavior(s) or outcomes were determined to be important, such as their level of difficulty, their frequency of performance, the consequences of error or other appropriate factors (essential). Where jobs are grouped or compared for the purposes of generalizing validity evidence, the work behavior(s) and work product(s) for each of the jobs should be described, and conclusions concerning the similarity of the jobs in terms of observable work behaviors or work products should be made (essential).

(5) Job titles and codes.

It is desirable to provide the selection procedure user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from the United States Employment Service's dictionary of occupational titles.

(6) Selection procedure.

The selection procedure used as a measure of the construct should be completely and explicitly described or attached (essential). If commercially available selection procedures are used, they should be identified by title, form and publisher (essential). The research evidence of the relationship between the selection procedure and the construct, such as factor structure, should be included (essential). Measures of central tendency, variability and reliability of the selection procedure should be provided (essential). Whenever feasible, these measures should be

provided separately for each relevant race, sex and ethnic group.

(7) Relationship to job performance.

The criterion-related study(ies) and other empirical evidence of the relationship between the construct measured by the selection procedure and the related work behavior(s) for the job or jobs in question should be provided (essential). Documentation of the criterion-related study(ies) should satisfy the provisions of section 15B above or section 15E(1) below, except for studies conducted prior to the effective date of these guidelines (essential). Where a study pertains to a group of jobs, and, on the basis of the study, validity is asserted for a job in the group, the observed work behaviors and the observed work products for each of the jobs should be described (essential). Any other evidence used in determining whether the work behavior(s) in each of the jobs is the same should be fully described (essential).

(8) Alternative procedures investigated.

The alternative selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of the investigation, and the conclusions reached in light of the findings should be fully described (essential).

(9) Uses and applications.

The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other procedures in a battery) and available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use, and the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If weights are assigned to different parts of the selection procedure, these weights and the validity of the weighted composite should be reported (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential).

(10) Accuracy and completeness.

The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

(11) Source data.

Each user should maintain records showing all pertinent information relating to its study of construct validity.

(12) Contact person.

The name, mailing address, and telephone number of the individual who may be contacted for further information about the validity study should be provided (essential).

E. Evidence of validity from other studies.

When validity of a selection procedure is supported by studies not done by the user, the evidence from the original study or studies should be compiled in a manner similar to that required in the appropriate section of this section 15 above. In addition, the following evidence should be supplied:

(1) Evidence from criterion-related validity studies.

a. Job information. A description of the important job behavior(s) of the user's job and the basis on which the behaviors were determined to be important should be provided (essential). A full description of the basis for determining that these important work behaviors are the same as those of the job in the original study (or studies) should be provided (essential).

b. Relevance of criteria. A full description of the basis on which the criteria used in the original studies are determined to be relevant for the user should be provided (essential).

c. Other variables. The similarity of important applicant pool or sample characteristics reported in the original studies to those of the user should be described (essential). A description of the comparison between the race, sex and ethnic composition of the user's relevant labor market and the sample in the original validity studies should be provided (essential).

d. Use of the selection procedure. A full description should be provided showing that the use to be made of the selection procedure is consistent with the findings of the original validity studies (essential).

e. Bibliography.

A bibliography of reports of validity of the selection procedure for the job or jobs in question should be provided (essential). Where any of the studies included an investigation of test fairness, the results of this investigation should be provided (essential). Copies of reports published in journals that are not commonly available should be described in detail or attached (essential). Where a user is relying upon unpublished studies, a reasonable effort should be made to obtain these studies. If these unpublished studies are the sole source of validity evidence they should be described in detail or attached (essential). If these studies are not available, the name

and address of the source, an adequate abstract or summary of the validity study and data, and a contact person in the source organization should be provided (essential).
(2) Evidence from content validity studies. See section 14C(3) and section 15C above.
(3) Evidence from construct validity studies. See sections 14D(2) and 15D above.

F. Evidence of validity from cooperative studies.

Where a selection procedure has been validated through a cooperative study, evidence that the study satisfies the requirements of sections 7, 8 and 15E should be provided (essential).

G. Selection for higher level job.

If a selection procedure is used to evaluate candidates for jobs at a higher level than those for which they will initially be employed, the validity evidence should satisfy the documentation provisions of this section 15 for the higher level job or jobs, and in addition, the user should provide: (1) a description of the job progression structure, formal or informal; (2) the data showing how many employees progress to the higher level job and the length of time needed to make this progression; and (3) an identification of any anticipated changes in the higher level job. In addition, if the test measures a knowledge, skill or ability, the user should provide evidence that the knowledge, skill or ability is required for the higher level job and the basis for the conclusion that the knowledge, skill or ability is not expected to develop from the training or experience on the job.

H. Interim use of selection procedures.

If a selection procedure is being used on an interim basis because the procedure is not fully supported by the required evidence of validity, the user should maintain and have available (1) substantial evidence of validity for the procedure, and (2) a report showing the date on which the study to gather the additional evidence commenced, the estimated completion date of the study, and a description of the data to be collected (essential).

(Approved by the Office of Management and Budget under control number 3046—0017)
(Pub. L. 96—511, 94 Stat. 2812 (44 U.S.C. 3501 et seq.))
[43 FR 38295, 38312, Aug. 25, 1978, as amended at 46 FR 63268, Dec. 31, 1981]
Definitions

§§1607.16 Definitions.
The following definitions shall apply throughout these guidelines:

A. Ability.

A present competence to perform an observable behavior or a behavior which results in an observable product.

B. Adverse impact.

A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group. See section 4 of these guidelines.

C. Compliance with these guidelines.

Use of a selection procedure is in compliance with these guidelines if such use has been validated in accord with these guidelines (as defined below), or if such use does not result in adverse impact on any race, sex, or ethnic group (see section 4, above), or, in unusual circumstances, if use of the procedure is otherwise justified in accord with Federal law. See section 6B, above.

D. Content validity.

Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job. See section 5B and section 14C.

E. Construct validity.

Demonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important for successful job performance. See section 5B and section 14D.

F. Criterion-related validity.

Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior. See sections 5B and 14B.

G. Employer.

Any employer subject to the provisions of the Civil Rights Act of 1964, as amended, including State or local governments and any Federal agency subject to the provisions of section 717 of the Civil Rights Act of 1964, as amended, and any Federal contractor or subcontractor or federally assisted construction contractor or subcontactor covered by Executive Order 11246, as amended.

H. Employment agency.

Any employment agency subject to the provisions of the Civil Rights Act of 1964, as amended.

I. Enforcement action.

For the purposes of section 4 a proceeding by a Federal enforcement agency such as a lawsuit or

an administrative proceeding leading to debarment from or withholding, suspension, or termination of Federal Government contracts or the suspension or withholding of Federal Government funds; but not a finding of reasonable cause or a conciliation process or the issuance of right to sue letters under title VII or under Executive Order 11246 where such finding, conciliation, or issuance of notice of right to sue is based upon an individual complaint.

J. Enforcement agency.

Any agency of the executive branch of the Federal Government which adopts these guidelines for purposes of the enforcement of the equal employment opportunity laws or which has responsibility for securing compliance with them.

K. Job analysis.

A detailed statement of work behaviors and other information relevant to the job.

L. Job description.

A general statement of job duties and responsibilities.

M. Knowledge.

A body of information applied directly to the performance of a function.

N. Labor organization.

Any labor organization subject to the provisions of the Civil Rights Act of 1964, as amended, and any committee subject thereto controlling apprenticeship or other training.

O. Observable.

Able to be seen, heard, or otherwise perceived by a person other than the person performing the action.

P. Race, sex, or ethnic group.

Any group of persons identifiable on the grounds of race, color, religion, sex, or national origin.

Q. Selection procedure.

Any measure, combination of measures, or procedure used as a basis for any employment decision. Selection procedures include the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs, or probationary periods and

physical, educational, and work experience requirements through informal or casual interviews and unscored application forms.

R. Selection rate.

The proportion of applicants or candidates who are hired, promoted, or otherwise selected.

S. Should.

The term ""should"" as used in these guidelines is intended to connote action which is necessary to achieve compliance with the guidelines, while recognizing that there are circumstances where alternative courses of action are open to users.

T. Skill.

A present, observable competence to perform a learned psychomotor act.

U. Technical feasibility.

The existence of conditions permitting the conduct of meaningful criterion-related validity studies. These conditions include: (1) An adequate sample of persons available for the study to achieve findings of statistical significance; (2) having or being able to obtain a sufficient range of scores on the selection procedure and job performance measures to produce validity results which can be expected to be representative of the results if the ranges normally expected were utilized; and (3) having or being able to devise unbiased, reliable and relevant measures of job performance or other criteria of employee adequacy. See section 14B(2). With respect to investigation of possible unfairness, the same considerations are applicable to each group for which the study is made. See section 14B(8).

V. Unfairness of selection procedure.

A condition in which members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences are not reflected in differences in measures of job performance. See section 14B(7).

W. User.

Any employer, labor organization, employment agency, or licensing or certification board, to the extent it may be covered by Federal equal employment opportunity law, which uses a selection procedure as a basis for any employment decision. Whenever an employer, labor organization, or employment agency is required by law to restrict recruitment for any occupation to those applicants who have met licensing or certification requirements, the licensing or certifying authority to the extent it may be covered by Federal equal employment opportunity law will be

considered the user with respect to those licensing or certification requirements. Whenever a State employment agency or service does no more than administer or monitor a procedure as permitted by Department of Labor regulations, and does so without making referrals or taking any other action on the basis of the results, the State employment agency will not be deemed to be a user.

X. Validated in accord with these guidelines or properly validated.

A demonstration that one or more validity study or studies meeting the standards of these guidelines has been conducted, including investigation and, where appropriate, use of suitable alternative selection procedures as contemplated by section 3B, and has produced evidence of validity sufficient to warrant use of the procedure for the intended purpose under the standards of these guidelines.

Y. Work behavior.

An activity performed to achieve the objectives of the job. Work behaviors involve observable (physical) components and unobservable (mental) components. A work behavior consists of the performance of one or more tasks. Knowledges, skills, and abilities are not behaviors, although they may be applied in work behaviors.

Appendix

§§1607.17 Policy statement on affirmative action (see section 13B).

The Equal Employment Opportunity Coordinating Council was established by act of Congress in 1972, and charged with responsibility for developing and implementing agreements and policies designed, among other things, to eliminate conflict and inconsistency among the agencies of the Federal Government responsible for administering Federal law prohibiting discrimination on grounds of race, color, sex, religion, and national origin. This statement is issued as an initial response to the requests of a number of State and local officials for clarification of the Government's policies concerning the role of affirmative action in the overall equal employment opportunity program. While the Coordinating Council's adoption of this statement expresses only the views of the signatory agencies concerning this important subject, the principles set forth below should serve as policy guidance for other Federal agencies as well.

(1) Equal employment opportunity is the law of the land. In the public sector of our society this means that all persons, regardless of race, color, religion, sex, or national origin shall have equal access to positions in the public service limited only by their ability to do the job. There is ample evidence in all sectors of our society that such equal access frequently has been denied to members of certain groups because of their sex, racial, or ethnic characteristics. The remedy for such past and present discrimination is twofold.

On the one hand, vigorous enforcement of the laws against discrimination is essential. But equally, and perhaps even more important are affirmative, voluntary efforts on the part of public employers to assure that positions in the public service are genuinely and equally accessible to qualified persons, without regard to their sex, racial, or ethnic characteristics. Without such efforts equal employment opportunity is no more than a wish. The importance of voluntary affirmative action on the part of employers is underscored by title VII of the Civil Rights Act of 1964, Executive Order 11246, and related laws and regulations——all of which emphasize voluntary action to achieve equal employment opportunity.

As with most management objectives, a systematic plan based on sound organizational analysis and problem identification is crucial to the accomplishment of affirmative action objectives. For this reason, the Council urges all State and local governments to develop and implement results oriented affirmative action plans which deal with the problems so identified.
The following paragraphs are intended to assist State and local governments by illustrating the kinds of analyses and activities which may be appropriate for a public employer's voluntary affirmative action plan. This statement does not address remedies imposed after a finding of unlawful discrimination.

(2) Voluntary affirmative action to assure equal employment opportunity is appropriate at any stage of the employment process. The first step in the construction of any affirmative action plan should be an analysis of the employer's work force to determine whether percentages of sex, race, or ethnic groups in individual job classifications are substantially similar to the percentages of those groups available in the relevant job market who possess the basic job-related qualifications.

When substantial disparities are found through such analyses, each element of the overall selection process should be examined to determine which elements operate to exclude persons on the basis of sex, race, or ethnic group. Such elements include, but are not limited to, recruitment, testing, ranking certification, interview, recommendations for selection, hiring, promotion, etc. The examination of each element of the selection process should at a minimum include a determination of its validity in predicting job performance.

(3) When an employer has reason to believe that its selection procedures have the exclusionary effect described in paragraph 2 above, it should initiate affirmative steps to remedy the situation. Such steps, which in design and execution may be race, color, sex, or ethnic ""conscious,"" include, but are not limited to, the following:

(a) The establishment of a long-term goal, and short-range, interim goals and timetables for the specific job classifications, all of which should take into account the availability of basically qualified persons in the relevant job market;

(b) A recruitment program designed to attract qualified members of the group in question;

(c) A systematic effort to organize work and redesign jobs in ways that provide opportunities for persons lacking ""journeyman"" level knowledge or skills to enter and, with appropriate training, to progress in a career field;

(d) Revamping selection instruments or procedures which have not yet been validated in order to reduce or eliminate exclusionary effects on particular groups in particular job classifications;

(e) The initiation of measures designed to assure that members of the affected group who are qualified to perform the job are included within the pool of persons from which the selecting official makes the selection;

(f) A systematic effort to provide career advancement training, both classroom and on-the-job, to employees locked into dead end jobs; and

(g) The establishment of a system for regularly monitoring the effectiveness of the particular affirmative action program, and procedures for making timely adjustments in this program where effectiveness is not demonstrated.

(4) The goal of any affirmative action plan should be achievement of genuine equal employment opportunity for all qualified persons. Selection under such plans should be based upon the ability of the applicant(s) to do the work. Such plans should not require the selection of the unqualified, or the unneeded, nor should they require the selection of persons on the basis of race, color, sex, religion, or national origin. Moreover, while the Council believes that this statement should serve to assist State and local employers, as well as Federal agencies, it recognizes that affirmative action cannot be viewed as a standardized program which must be accomplished in the same way at all times in all places.

Accordingly, the Council has not attempted to set forth here either the minimum or maximum voluntary steps that employers may take to deal with their respective situations. Rather, the Council recognizes that under applicable authorities, State and local employers have flexibility to formulate affirmative action plans that are best suited to their particular situations. In this manner, the Council believes that affirmative action programs will best serve the goal of equal employment opportunity.

Respectfully submitted,

Harold R. Tyler, Jr.,
Deputy Attorney General and Chairman of the Equal Employment Coordinating Council.

Michael H. Moskow,
Under Secretary of Labor.

Ethel Bent Walsh,
Acting Chairman, Equal Employment Opportunity Commission.

Robert E. Hampton,
Chairman, Civil Service Commission.

Arthur E. Flemming,
Chairman, Commission on Civil Rights.

Because of its equal employment opportunity responsibilities under the State and Local Government Fiscal Assistance Act of 1972 (the revenue sharing act), the Department of Treasury was invited to participate in the formulation of this policy statement; and it concurs and joins in the adoption of this policy statement.
Done this 26th day of August 1976.
Richard Albrecht,
General Counsel,
Department of the Treasury.

§§1607.18 Citations.
The official title of these guidelines is ""Uniform Guidelines on Employee Selection Procedures (1978)"". The Uniform Guidelines on Employee Selection Procedures (1978) are intended to establish a uniform Federal position in the area of prohibiting discrimination in employment practices on grounds of race, color, religion, sex, or national origin. These guidelines have been adopted by the Equal Employment Opportunity Commission, the Department of Labor, the Department of Justice, and the Civil Service Commission.

The official citation is:
Section &lowbarm;&lowbarm;, Uniform Guidelines on Employee Selection Procedure (1978); 43 FR &lowbarm;&lowbarm; (August 25, 1978).
The short form citation is:
Section &lowbarm;&lowbarm;, U.G.E.S.P. (1978); 43 FR &lowbarm;&lowbarm; (August 25, 1978).

When the guidelines are cited in connection with the activities of one of the issuing agencies, a specific citation to the regulations of that agency can be added at the end of the above citation.
The specific additional citations are as follows:
Equal Employment Opportunity Commission
29 CFR part 1607
Department of Labor
Office of Federal Contract Compliance Programs
41 CFR part 60—3
Department of Justice
28 CFR 50.14

Civil Service Commission
5 CFR 300.103(c)
Normally when citing these guidelines, the section number immediately preceding the title of the guidelines will be from these guidelines series 1—18. If a section number from the codification for an individual agency is needed it can also be added at the end of the agency citation. For example, section 6A of these guidelines could be cited for EEOC as follows:
Section 6A, Uniform Guidelines on Employee Selection Procedures (1978); 43 FR &lowbarm;&lowbarm;, (August 25, 1978); 29 CFR part 1607, section 6A.

****