

# Data Management Manual

National Center on  
Assessment and Accountability  
For Special Education

Version 3 – June 15, 2013

## **CHAPTER 1**

### **OVERVIEW OF NCAASE DATA MANAGEMENT**

Data management is critical, particularly in the context of multiple states (AZ, NC, OR, and PA), multiple types of state assessments (general and alternate), multiple years of test data (2004 – 2012), and multiple test vendors (Northwest Evaluation Association and Behavioral Research and Teaching – University of Oregon), and multiple researchers (four Principal Investigators and at least three associated faculty). The purpose of this document is to establish guidelines for such data management. Five broad issues are addressed that allow researchers to conduct analyses with clean data sets that accurately represent the variables and values needed to use specific analytic software. These issues do not, however, address any specific considerations for analyzing data (assumptions, settings, and procedures).

#### **NCAASE Datasets and Research Questions**

Four states are participating with this center to address a number of questions about student growth in reading and mathematics achievement: Arizona, North Carolina, Oregon, and Pennsylvania. The datasets from these states include both the statewide test given to all students (including those in special education) and the alternate assessment given to students with the most significant cognitive disabilities (1%). In addition, Northwest Evaluation Association (NWEA) is a participant with the Measure of Academic Performance (MAP).

Because the emphasis of this research is centered on growth, the data represent a number of successive years (though the beginning years vary for different states. Following is a list of the state and the years in which data are available with a column noting any obvious anomalies in the data. In the two tables below, the general test used in statewide accountability is labeled

GEN and the alternate assessment test used for 1% of the district population with significant cognitive disabilities is labeled AA.

*Table 1. Available state datasets available for NCAASE researchers*

State – GEN	Years	Notes
Arizona	2007-2011	Includes Alternate Assessment
North Carolina	2001-2011	Test edition changes made in reading and math
Oregon	2004-2012	2006-2007 is split with paper pencil and computer
Pennsylvania	2006-2012	Includes Alternate Assessment
NWEA		

State – AA	Years	Notes
Arizona	2007-2011	
North Carolina		With 2%; Alternate Assessment added in 2006
Oregon	2004-2012	2007-2008 first year with consistent task-items
Pennsylvania	2006-2012	
NWEA		

easyCBM data are also present for reading and mathematics with the following caveats:

- Grades K-8 contain specific measures that are listed in the appendices of Chapter 3.
- Benchmark and progress monitoring datasets have different data structures.
- Measurement development is ongoing so the specific year needs to be considered.

easyCBM data are downloaded from the Lite and District versions on a weekly basis with seasonal benchmark measures compiled into (a) all measures by CBM type and grade, (b)

benchmark item level files by grade, (c) interventions. Beginning in 2012-2013, roster uploads for the District versions include student demographics.

*Table 2. Available easyCBM datasets available for NCAASE researchers*

<b>Reading</b>	<b>Measure</b>	<b>Notes</b>
2010-2011	K-8 aligned with National Reading Panel	Primarily Oregon Benchmarks; limited demographics
2011-2012	K-8 aligned with National Reading Panel	National sample with full Oregon demographics
2012-2013	3-8 aligned with CCSS	National sample with full demographics

<b>Mathematics</b>	<b>Measure</b>	<b>Notes</b>
2010-2011	K-8 aligned with National Council of Teachers of Mathematics	Primarily Oregon Benchmarks; limited demographics
2011-2012	K-8 aligned with National Council of Teachers of Mathematics	National sample with (nearly) full Oregon demographics
2012-2013	3-8 aligned with CCSS	National sample with full demographics: (a) sex, (b) race, (c) ethnicity, (d) English Language learner status, and (e) special education services

*Table 3. Study focus, research questions and datasets addressed over five years*

Study	Primary Research Question(s) Addressed*	Datasets
Cornerstone Study Years 1-2, 5	1. What is the natural developmental progress in achievement for students with disabilities? 2. What models best characterize achievement growth for students with disabilities who are participating in general achievement tests? 3. How do various growth models represent school effects for students with and without disabilities, and how do results compare to those derived from status models now in use? 4. What are the reliability and validity of estimates of school effectiveness for students with disabilities produced by alternative growth models and how are these estimates influenced by contextual differences among schools and students?	NC Statewide Achievement Test (Grades 3-8: Reading & Math)
Multi-State Extension Study Years 2 & 5	<u>Across 3 partner states:</u> 1. What is the natural developmental progress in achievement for students with disabilities? 2. What models best characterize achievement growth for students with disabilities who are participating in general achievement tests? 3. How do various growth models represent school effects for students with and without disabilities, and how do results compare to those derived from status models now in use? 4. What are the reliability and validity of estimates of school effectiveness for students with disabilities produced by alternative growth models and how are these estimates influenced by contextual differences among schools and students?	AZ Statewide Achievement Test (Grades 3-8: Reading & Math)
		OR Statewide Achievement Test (Grades 3-8: Reading & Math)
		PA Statewide Achievement Test (Grades 3-8: Reading & Math)
Interim Assessments Study Years 1 & 2	<u>For 2 interim assessments:</u> 1. What is the natural developmental progress in achievement for students with disabilities? 5. How do results from different types of interim assessments of students' achievement meaningfully contribute to a model of academic growth for students with disabilities?	NWEA MAP in reading and mathematics Grades 1-10  easyCBM measures in reading and mathematics Grades 1-8
Multiple Measures Validation Study Years 2-4	5. How do results from different types of interim assessments of students' achievement meaningfully contribute to a model of academic growth for students with disabilities? 6. How can information about opportunity to learn and achievement growth be used to enhance academic outcomes for students with disabilities?	Multiple Measures OTL+ Interim Assessments [CBM + Multiple Choice Tests] + OR & PA Summative Assessments (Grades 3-5: Reading & Math)
Alternate Assessments Study Years 3-5	1. What is the natural developmental progress in achievement for students with disabilities? 2. What models best characterize achievement growth for students who are participating in alternate assessments?	Alternate Assessment (Grades 3-8 for Reading & Math): AZ • NC• OR • PA

**Transferring, storing, and accessing data sets** consistent with best security practice is the first step in the data management process. In **Chapter 2** the process is outlined in which original files are accessed in a secure manner from research partners and stored in a password protected University of Oregon (UO) server; the original files are stored in a volume only accessible by the Principal Investigators or BRT faculty. Data sets are then transferred to a broader group of NCAASE researchers via one of three methods: (a) using a secure data transport to off campus researchers (coe-ncaase.stats.uoregon.edu), (b) moving files directly to another UO server with statistical software available, or (c) through web-based data ordering (that then employs the data transport system). *All original and final files reside in datastore*, which is a BRT server hosted by the University of Oregon with state data stored in partitions available only to NCAASE researchers.

With data sets securely available, **understanding the data directories** is the next critical issue: variables and descriptions of what they represent and how they are coded. The data are measure-specific and though they often reflect common data types, the properties and coding values differ. These issues are addressed in **Chapter 3** with references to an extensive number of appendices for the measures used in the NCAASE research.

A number of issues are addressed in **processing data sets**, including data file codebooks that document file structures (variable names, descriptors, and value ranges), data preparation (elimination of duplicate records and resolution of data discrepancies such as out of range values and cross-variable inconsistencies), and establishment of longitudinal files (concatenation of successive annual files and resolution of missing data or individual file

inconsistencies). These issues are addressed in **Chapter 4** including a general description of critical components and, in a series of appendices, specific data set documentation.

The final step in data management involves **structuring specific data sets** for analysis. We use two methods for organizing this process: (a) a remote desktop connection in which files can be manipulated, and (b) direct access to a database in which queries are used to export files. Both approaches use a structured query language (SQL) for manipulating data sets to be consonant with the critical questions being addressed through further statistical analyses. Both systems are described in **Chapter 5**.

This **Data Management Manual** can serve as a resource for NCAASE researchers to access and analyze high quality data in addressing critical issues in growth modeling. General guidelines and specific procedures allow efficient use of protocols in getting the data sets into the proper form and format. Annual revisions are planned as the datasets sets change.

## CHAPTER 2

### SECURITY, ACCESS, TRANSFER, AND STORAGE OF DATASETS

The purpose of this chapter is to describe the process for researchers to access and transfer data from research projects being conducted by Behavioral Research and Teaching (BRT) under either the auspices of the National Center on Assessment and Accountability for Special Education (NCAASE) or eacyCBM (District version). Following is a conceptual rendering of the datasets and transfer and then an articulation of the actual steps to for accessing and sharing files. The conceptual rendering is important primarily in orientating and understanding the ‘big ticket issues’. The specific protocols that can be referenced for access to the NCAASE stats server and to transfer files.

#### **Data Security**

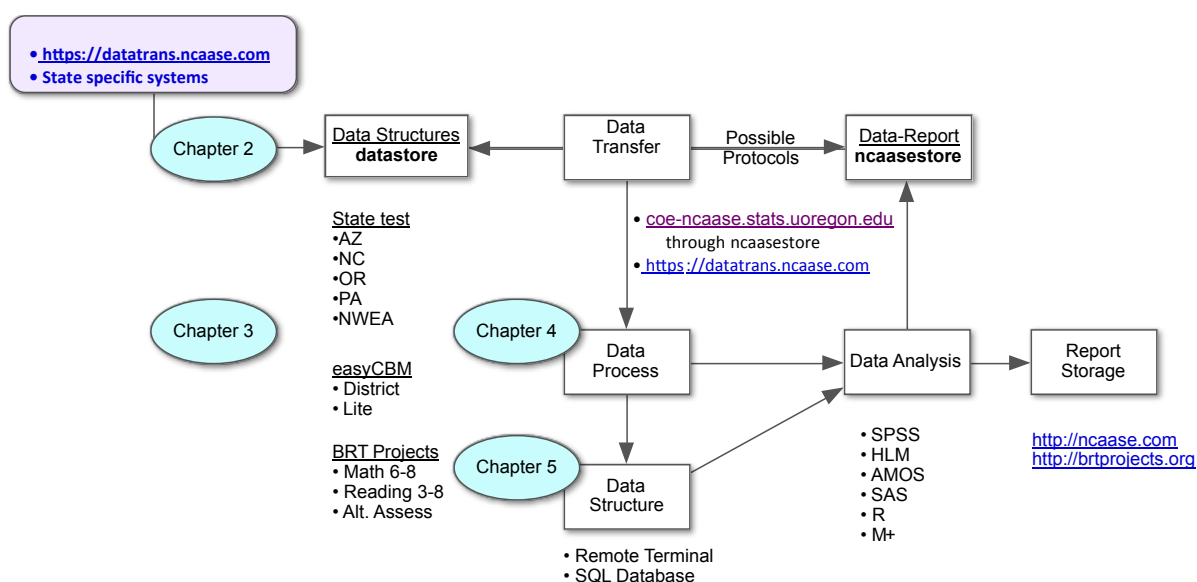
NCAASE has adopted several policies on data security, all of which govern access, transfer, and storage of data that have been developed internally or in association with various agencies or institutions: (a) Institutional Review Board (appendix 2.1), (b) North Carolina document on access and storage, (c) BRT computer security policy, (d) North Carolina data use agreement, (e) multi-institution generic data use agreement applicable for access to data from other states, and (f) internal NCAASE study proposal form.

*<See Appendix 2.1 to 2.6>*

In addition, data backup is comprised of the following safeguards. All dataset storage and server based processing is performed on equipment designed for professional use including redundancies in common points of failure such as hard disks and power supplies. The

computers and servers are physically protected with either mechanical or electronic locks and only essential personnel have access, all of them University of Oregon employees, affiliated and approved faculty, and doctoral students. Data files themselves are protected against unwanted change and deletion, and for future auditing purposes by a series of off-machine backups. The backups copy complete files and/or changes to the files and are stored on similarly secure infrastructure. Most of the research is conducted through personal computers using a virtual computing environment (with the actual data sets and files located on a secure UO server).

### Data Management System



### NCAASE Researchers with Access

Three groups of researchers are designed to have controlled access to datasets.

1. NCAASE Principal Investigators – In the grant application, the following PIs (and their respective institutions) have access to all state data sets with approval accompanied by the original IRB (Protocol No: 07052011.005): Ann Schulte (NCSU and ASU), Steve Elliott (ASU), Joe

Stevens, Joe Nese, and Gerald Tindal (UO).

2. NCAASE Affiliated Research Faculty – Three types of individuals are associated with this group: grant-funded researchers, advisors, and consultants. Presently, only grant-funded researchers have access to the data files: Keith Zvoch and Gina Biancarosa (UO) as well as Alex Kurz and Roy Levy (ASU). As (if) either advisors or consultants become involved in data analysis, a de-identified file needs to be created for any files that are transferred off the UO secure servers.

3. Doctoral students working with either PIs or Affiliated faculty. Presently, this group works in the virtual computing environment under the auspices of key NCAASE PIs.

Each researcher from these three groups needs to have a folder on the NCAASE server where all the structured files are located. Once provided a secure folder, researchers can become an established virtual user on the server and access files in a protected folder. After being added to the server (with a folder), they need to log in to establish their connectivity. Later in this chapter, steps are provided for transferring and deleting files being used in analyses.

### **Original Files in ncaasestore**

All original files are located in *ncaasestore*, a BRT server share. It is accessed only by the Principal Investigators. Once a file has been uploaded to *ncaasestore*, it can be moved to the NCAASE stats server (<rdp://coe-ncaase-stats.uoregon.edu>) so the files are accessible to other NCAASE researchers (through the Data Manager, Dr. Nese, or the Director of BRT, Dr. Tindal serving as a backup). The data manager with NCAASE moves a state file into a public share folder on the *coe-ncaase-stats* server. At that point, any of the approved researchers can access

the file from public shared folder and place it in their own secure folder. This file is then deleted from the public folder. This public folder is the default Windows OS specific “public” folder, where every person with a user account on the computer shares this folder. There is no general access to this folder, only pre-authorized users with accounts on the machine.

### **Accessing NCAASE Stats Server ([coe-ncaase-stats.uoregon.edu](https://coe-ncaase-stats.uoregon.edu))**

Data should be processed prior to movement into researcher folders, as this is the step used to clean files and provide only ONE original file that may be migrated for individual structuring. Dr. Nese is the Data Manager to contact for transferring files into the BRT *ncaasestore* server and then moving them to the NCAASE stats server.

The NCAASE server has all state datasets located in the C drive. Several software applications also are available including SPSS, HLM7, Mplus, R, and SAS. Once logged into the NCAASE stats server, researchers can access a shared folder with data files that are nested as follows: *Users...Public... NCAASE-Shared-StatsFiles*. In addition, all users have the default and “home” directory structure that is protected from other system users.

### **File Transfer with datatrans**

A web application (<https://datatrans.ncaase.com>) is available for researchers to transfer files in a secure manner amongst each other. N.B. Security implications of using personal computers such as laptops should be considered before using this service – transfers from within the [coe-ncaase-stats.uoregon.edu](https://coe-ncaase-stats.uoregon.edu) server are most secure.

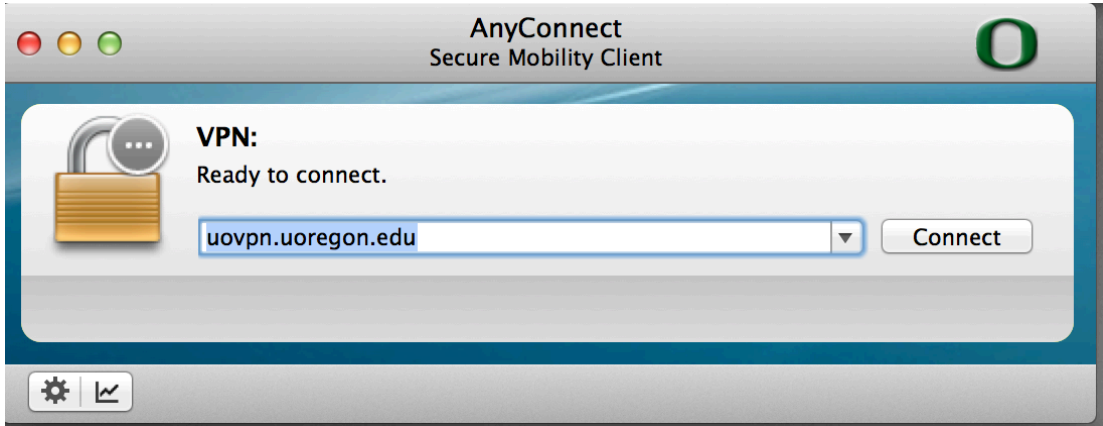
1. All NCAASE PIs can serve as administrators on the system. Not only can administrators establish users (create, delete, and change user passwords) but they also can monitor all transactions (e.g., see transfer status and history).

2. A user must have an account on the system to upload a file (e.g., have a username and pw).

3. Users log into the site, browse for a file to transfer, upload it, and receive a secure URL for transmission to the recipient. Advanced options are available in which the user types in the email of the person to receive the file, a password, and the number of days before automatic file deletion occurs.

4. To receive a file, the user accesses the unique file transfer URL (typical use case: sent by email) that may or may not also require a password (typical use case: sent via text or phone by person initiating the transfer).

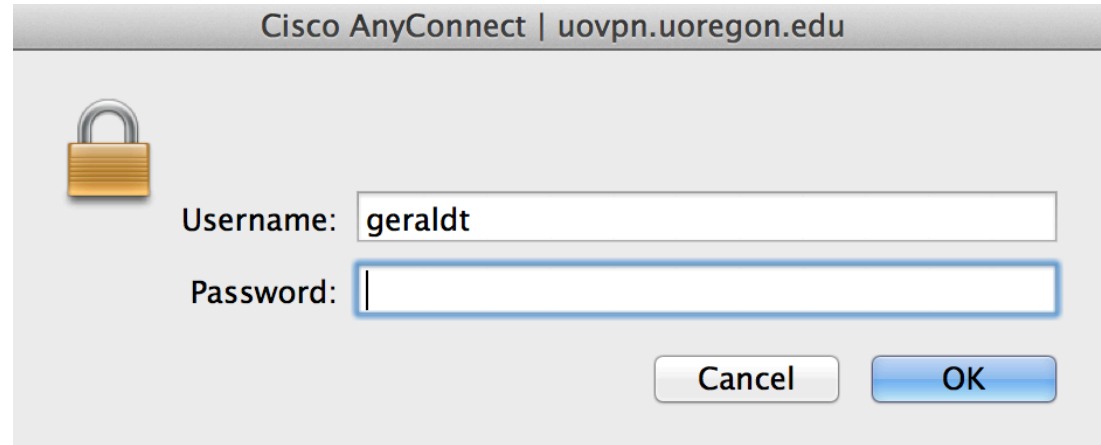
**Steps for Accessing NCAASE Stats Server**

<p>Screen 1:</p> <p>Cisco Mobility Client (for off campus)</p> <ul style="list-style-type: none"><li>• Remote Desktop Access to the NCAASE stats server is available from within the University of Oregon network only. If outside of the University of Oregon network, a VPN (Virtual Private Network) connection can first be established to “join” the University of Oregon network.</li></ul>	
<p>Screen 1b:</p> <ul style="list-style-type: none"><li>• Go to the following site for instructions to download the Cisco AnyConnect VPN Software.</li></ul>	<p><a href="https://it.uoregon.edu/vpn/anyconnect">https://it.uoregon.edu/vpn/anyconnect</a></p>

Screen 2:

Enter UO password


- Must have University of Oregon “duckid” in order to utilize UoFO VPN services.
- Additional information on duckid including ability to change password available at: <https://duckid.uoregon.edu>



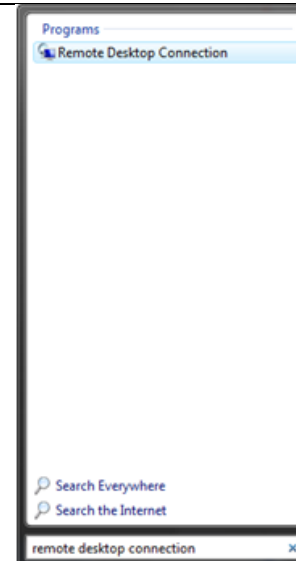
The image shows a Cisco AnyConnect login window titled "Cisco AnyConnect | uovpn.uoregon.edu". On the left is a yellow padlock icon. To its right are two input fields: "Username:" with the text "geraldt" and "Password:" which is empty. Below the fields are two buttons: "Cancel" and "OK".

Screen 3b:

- PC: Click on the **Start** button to open your

Start Menu . In the search bar, type *remote desktop connection* and open the program.

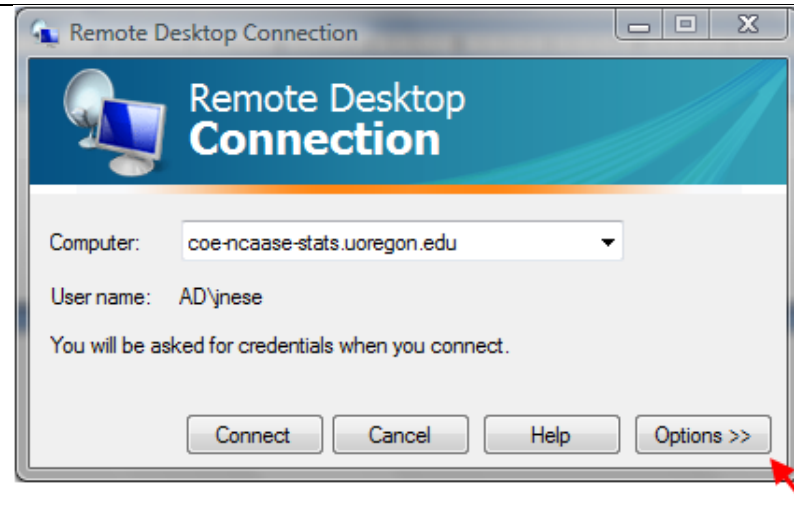
- Type in the name of the computer: *coe-ncaase-stats.uoregon.edu*



Screen 3c:

PC: *Options* button

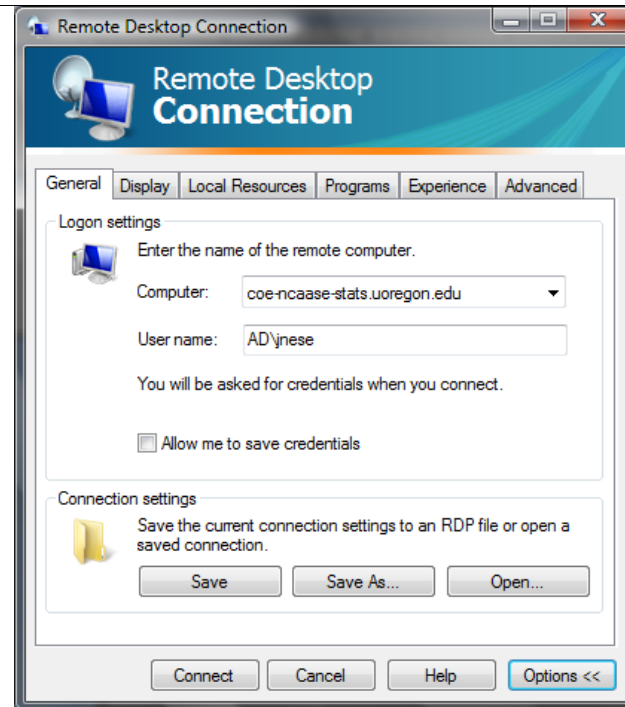
- Press the options button.

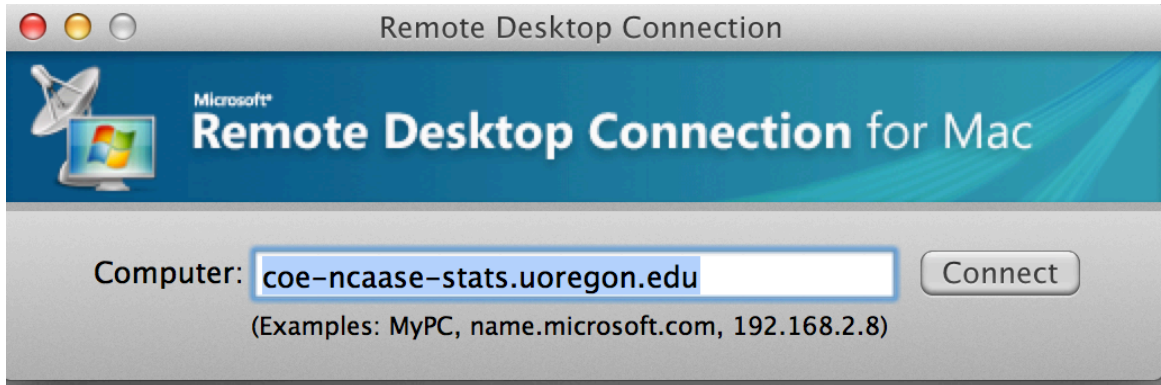



Screen 3d:

PC: User name:

- Your user name will be AD\your Duck ID (e.g., AD\jnese)



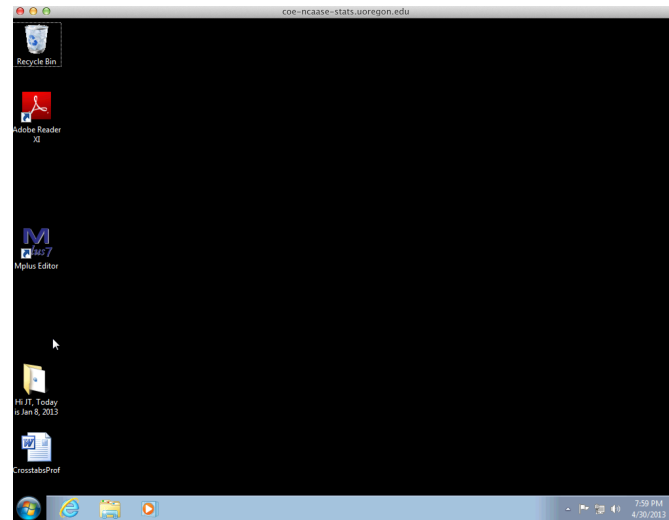
<p>Screen 3e:</p> <p>Remote Desktop Connection</p> <ul style="list-style-type: none"><li>• Mac: To establish connection with NCAASE stats server, users must have a Remote Desktop Connection client. Free clients exist for most popular operating systems.</li></ul>	
<p>Screen 4:</p> <p>Mac: Press Connect</p> <ul style="list-style-type: none"><li>• This message is presented when the optional “Mac OS Drives” preference pane has been configured in the Remote Desktop Client. Depending on individual configuration, users may not see this message.</li></ul> <p>N. B. This message is due to an optional preference set in the Mac OS Remote Desktop Client preferences pane (see pic below). It’s not the default so many users will never see this message.</p>	

Screen 5:  
Press Connect

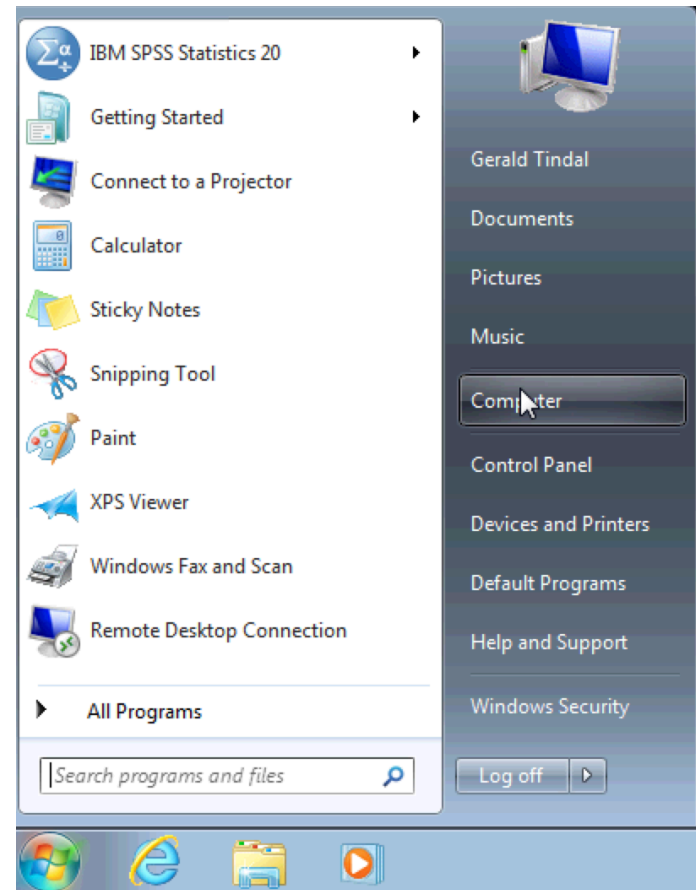
N.B. This message comes from using UO  
servers which are by default secure



Screen 6:  
View of the Server

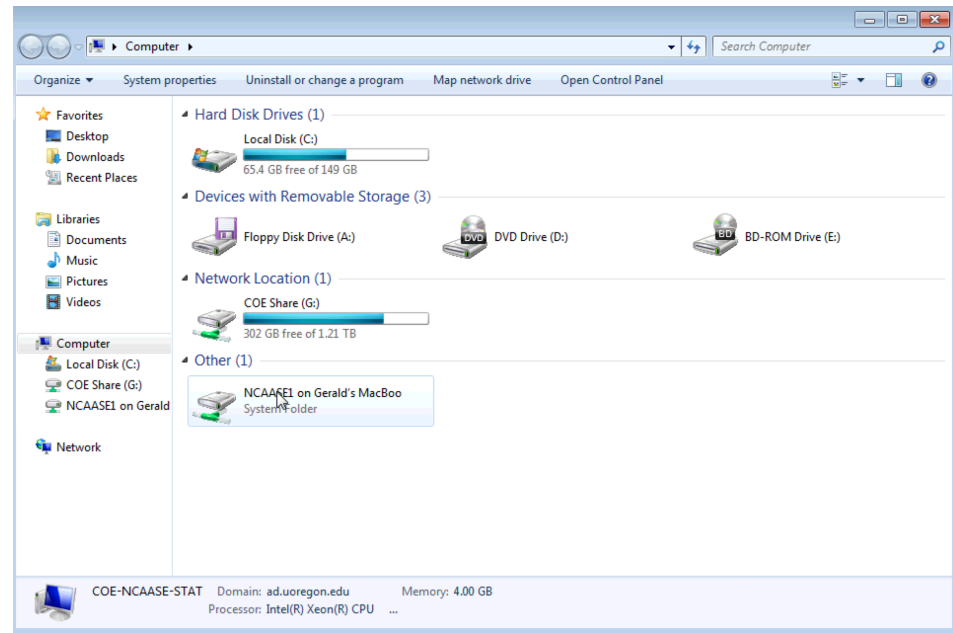


Screen 7:  
Browse Computer for access to Folders and  
Files



## Screen 8:

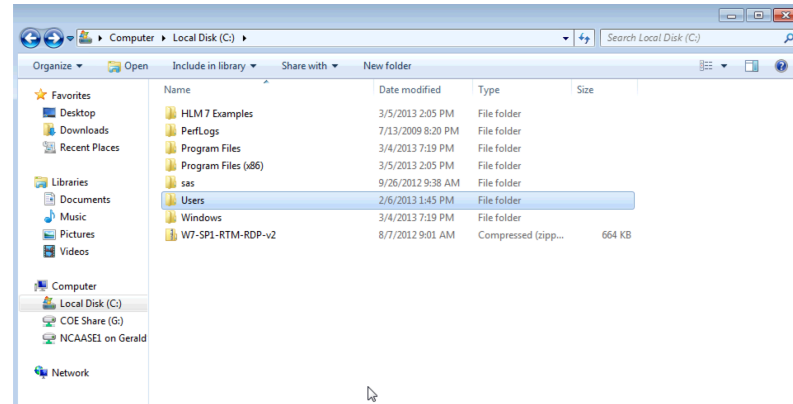
Windows browser displaying local disks, attached shares, and available shares.



## Screen 9:

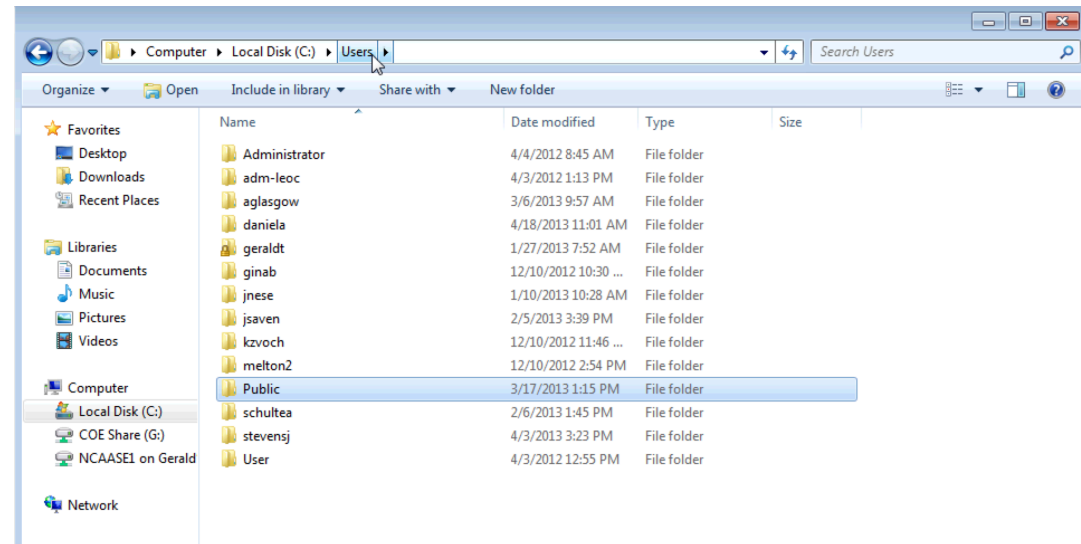
## Select Users folder

- Users folder exists on the C: drive

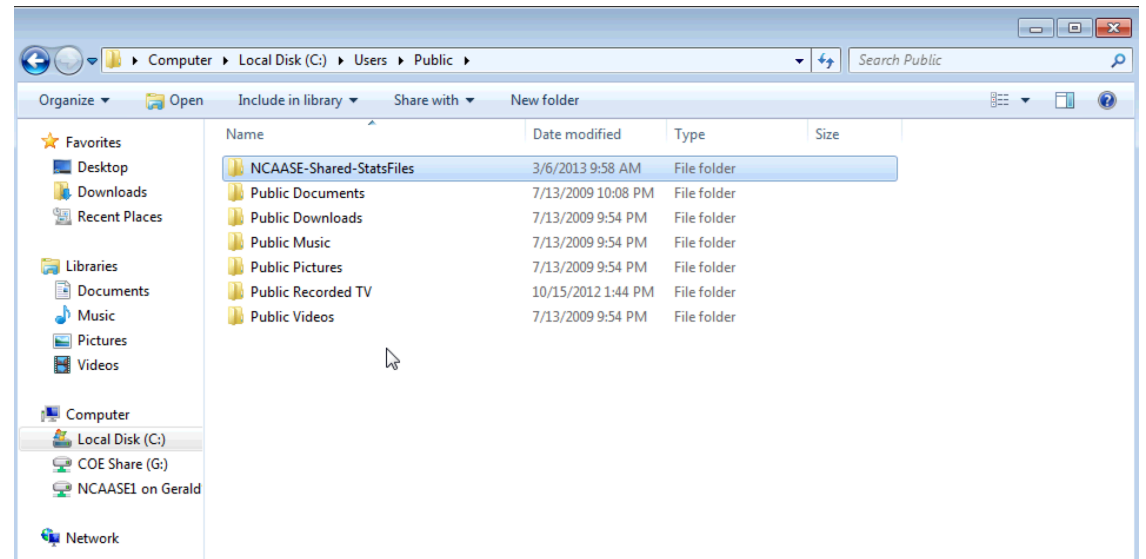


## Screen 10:

Public folder is the interchange for passing files back and forth among users with access to this machine.

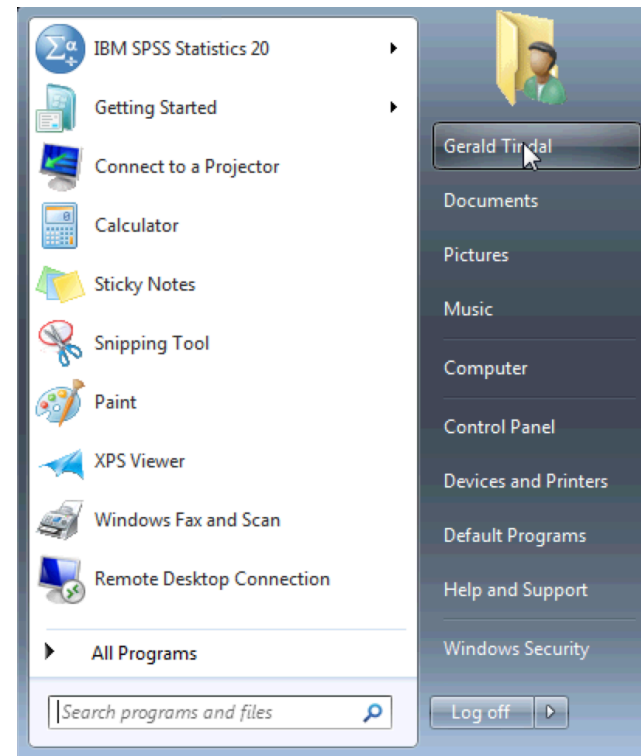


Screen 11:  
Inside Public is NCAASE-Shared-StatsFiles



## Screen 12:

Instead of navigating through computer to Public to NCAASE-Shared-StatsFiles...Go directly to personal folder (Gerald Tindal is mine).

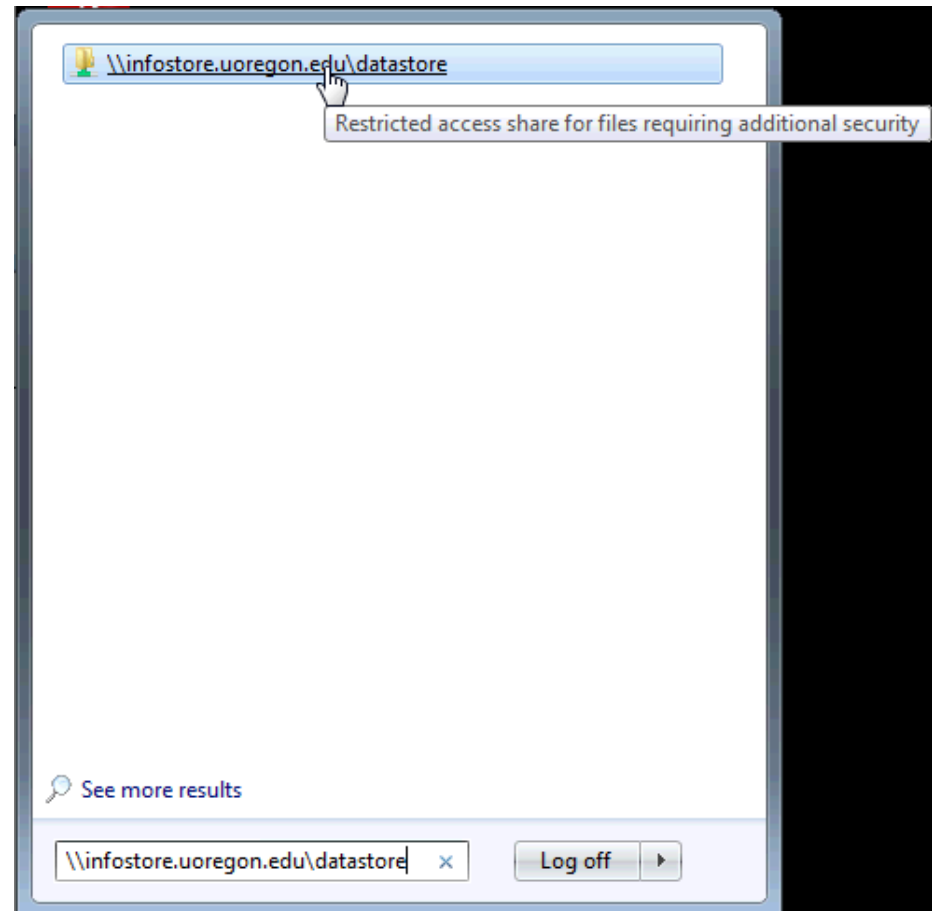


## Screen 13:

To mount the CIFS share datastore (for transferring files), click on Microsoft logo (lower left corner of Screen 7). Type:

[\\infostore.uoregon.edu\datastore](https://infostore.uoregon.edu/datastore)

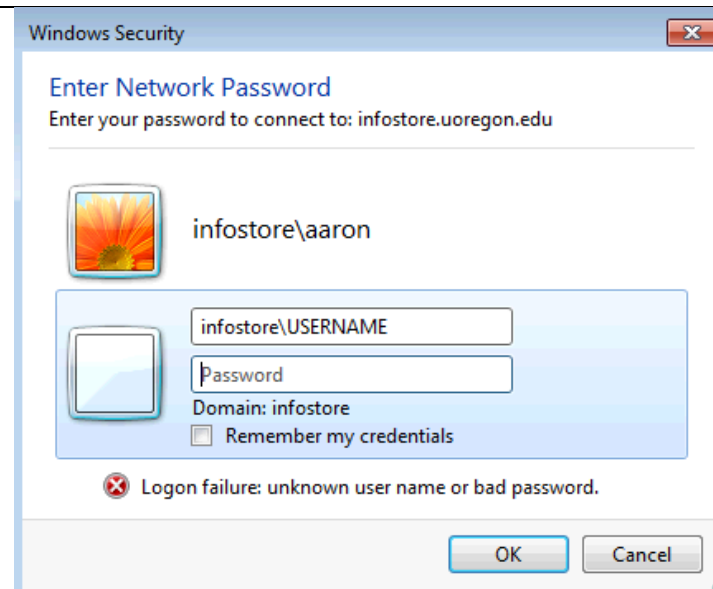
- datastore and ncaase store available to users with additional access privileges only. Currently these are UofO based PIs.



## Screen 14:

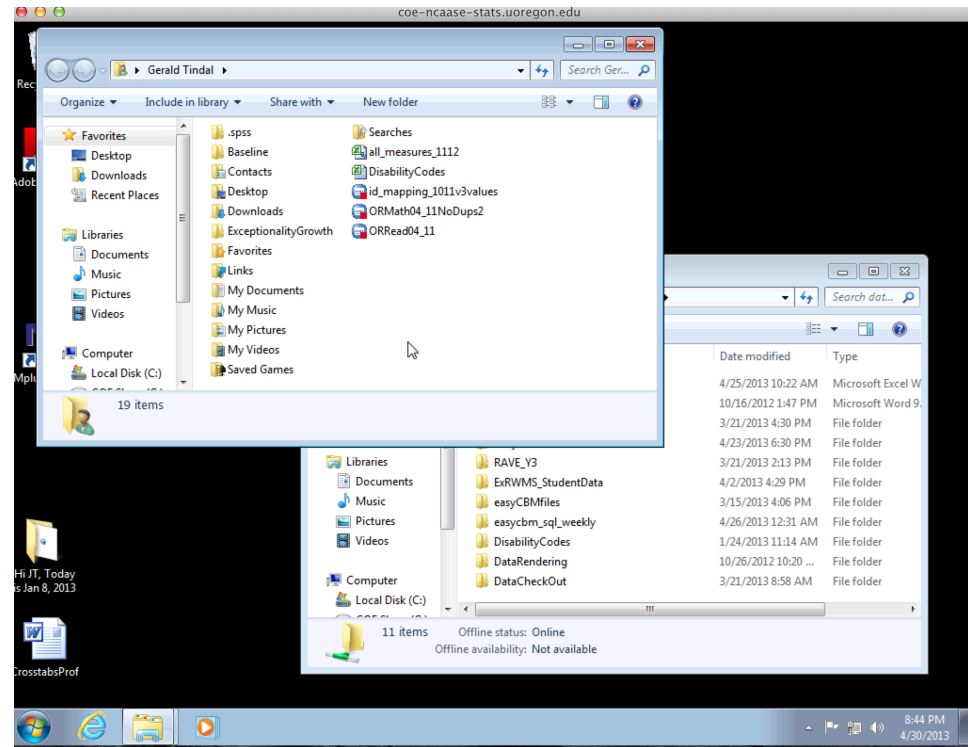
Enter your credentials for the BRT administered infostore server. This server provides the shares datastore and ncaasestore.

- The infostore server does not use the default Active Directory authentication. Precede your username with infostore\ to indicate authentication using an account that's local to the server.



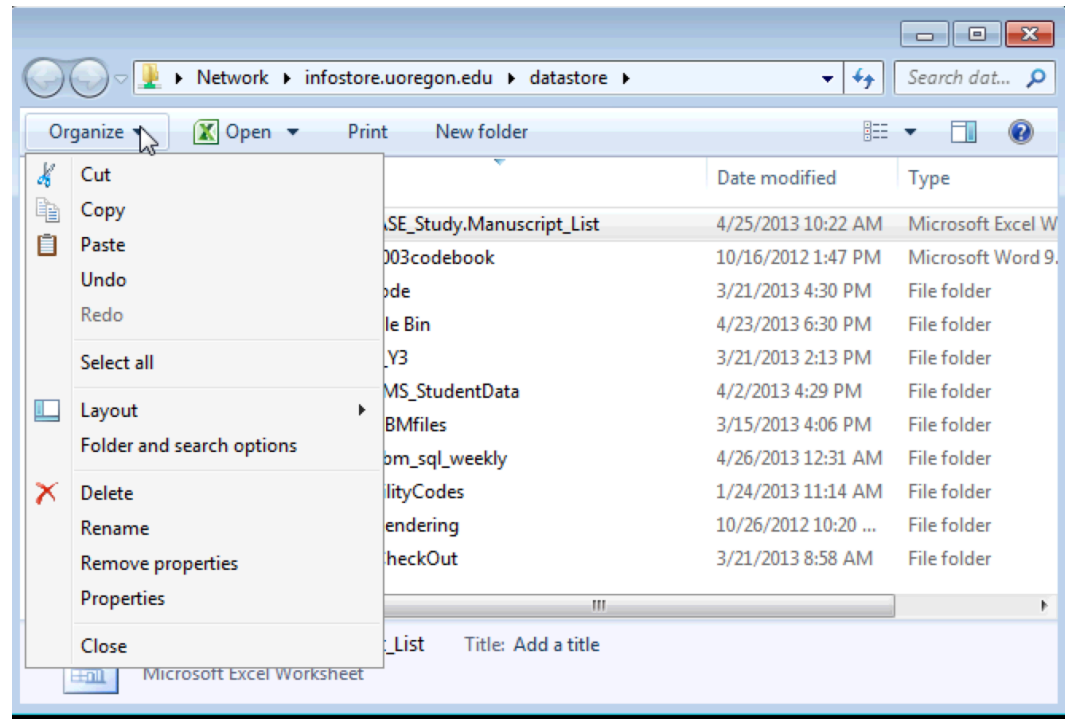
## Screen 15:

Once datastore is mounted, files can be copied and pasted (moved) to and from coe-ncaase-stats (use Organize on upper left banner)

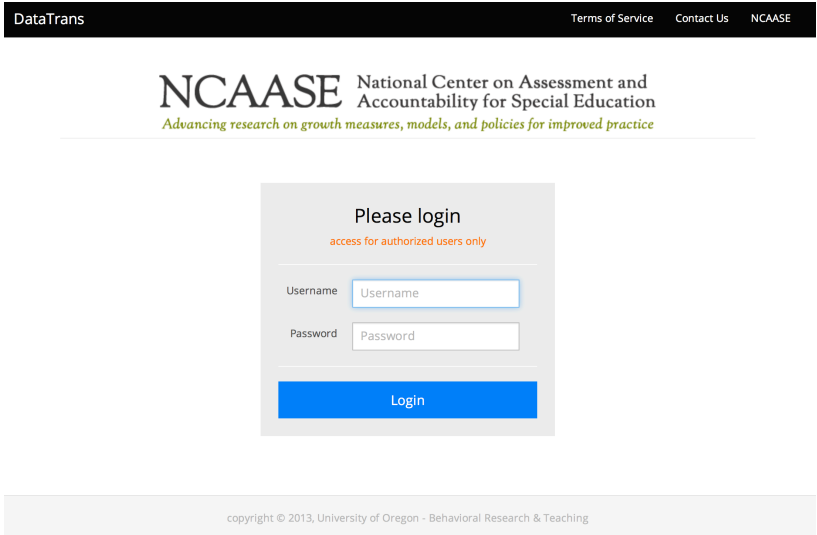
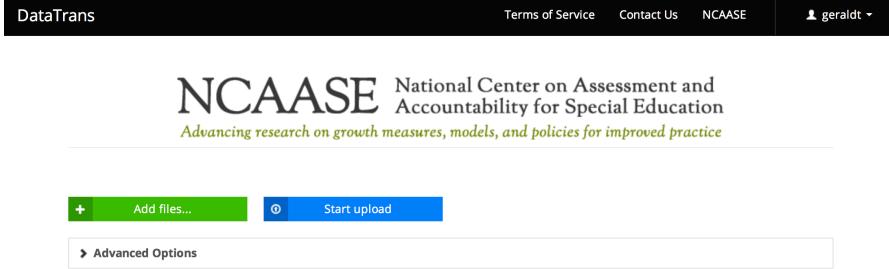
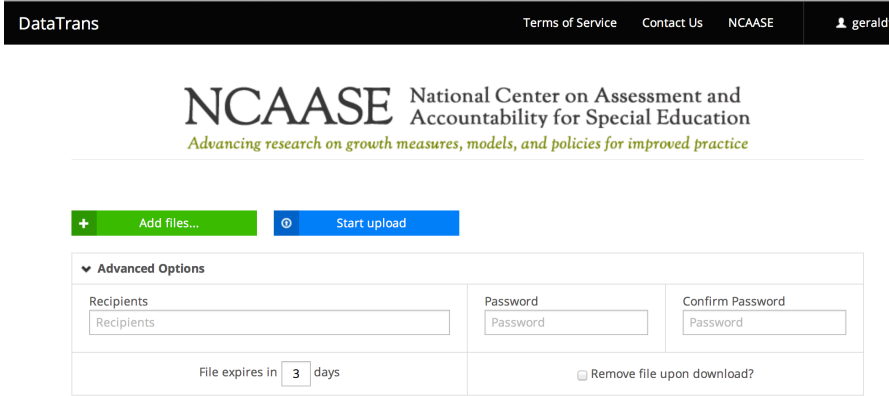


Screen 16:

Click on Organize to view file manipulation options (cut, copy, paste)



## Steps for Transferring Files via <http://datatrans.ncaase.com>

<p>Screen 1:</p> <p>Using a web browser, navigate to:  <a href="https://datatrans.ncaase.com">https://datatrans.ncaase.com</a></p> <ul style="list-style-type: none"> <li>• Must have an account to user this service, contact NCAASE data manager: Joe Nese, <a href="mailto:jnese@uoregon.edu">jnese@uoregon.edu</a></li> </ul>	
<p>Screen 2:</p> <p>Open page to upload a file</p>	
<p>Screen 3:</p> <p>Add security options of user name and password</p>	

Screen 4:  
Unique file transfer link,  
to be provided to  
individual(s) whom you  
wish to transfer file to.

- If advanced security options are selected such as an additional password, this information must also be transferred to the receiving individual.

## NCAASE National Center on Assessment and Accountability for Special Education

*Advancing research on growth measures, models, and policies for improved practice*

**Success:** All files have been uploaded and are now available for download via the link below!

File	Size	Expiration
HCTA review Tindal 051513.docx	118.14 KB	May 31st, 2013 1:51 PM

### Download Link

<https://datatrans.ncaase.com/download/?key=35e55f75fc64eabbef2e9abc4c3539787a9fcb904a4526f42ae06c2253c57be4>

Screen 5:  
View for recipient to  
download a file (once  
they have clicked on the  
link)

DataTrans

[Terms of Service](#)

[Contact Us](#)

[NCAASE](#)

## NCAASE National Center on Assessment and Accountability for Special Education

*Advancing research on growth measures, models, and policies for improved practice*

File	Size	Expiration	
HCTA review Tindal 051513.docx	118.14 KB	May 31st, 2013 1:51 PM	<a href="#">Download</a>

Screen 6:  
Administrator  
view of users  
(and to create  
or edit users)

- Only available to system administrators
- New users are sent an email with their default password and can then log in to change this password

DataTrans Terms of Service Contact Us NCAASE geraldtd

**NCAASE** National Center on Assessment and Accountability for Special Education  
*Advancing research on growth measures, models, and policies for improved practice*

Users Uploads Downloads Expired

Username	Email Address	Full Name	Is Active?	Is Admin?	+	Create New User
tcords	<a href="mailto:tcords@uoregon.edu">tcords@uoregon.edu</a>	Trevor Cords	Yes	Yes		Edit Delete
jnese	<a href="mailto:jnese@uoregon.edu">jnese@uoregon.edu</a>	Joseph Nese	Yes	Yes		Edit Delete
geraldtd	<a href="mailto:geraldtd@uoregon.edu">geraldtd@uoregon.edu</a>	Gerald Tindal	Yes	Yes		Edit Delete
aglasgow	<a href="mailto:aaron.glasgow@gmail.com">aaron.glasgow@gmail.com</a>	Aaron Glasgow	Yes	Yes		Edit Delete
aarong	<a href="mailto:aglasgow@gmail.com">aglasgow@gmail.com</a>	Aaron Glasgow	Yes	No		Edit Delete
jerryt	<a href="mailto:jerry@dillardresearchassociates.com">jerry@dillardresearchassociates.com</a>	jerry tindal	Yes	No		Edit Delete
DJAnderson07	<a href="mailto:daniela@uoregon.edu">daniela@uoregon.edu</a>	daniela@uoregon.edu Anderson	Yes	No		Edit Delete

Screen 7:  
Administrator  
view of uploads

DataTrans Terms of Service Contact Us NCAASE geraldtd

**NCAASE** National Center on Assessment and Accountability for Special Education  
*Advancing research on growth measures, models, and policies for improved practice*

Users Uploads Downloads Expired

Uploaded By	Name	Size	Link	Date Uploaded	Date Expires	Is Nonce?	Has Password?	Recipients
Gerald Tindal	HCTA review Tindal 051513.docx	118.14 KB	<a href="#">Download</a>	05-28-2013 1:51 PM	05-31-2013 1:51 PM	No	No	
jerry tindal	HCTA review Tindal 051513.docx	118.14 KB	<a href="#">Download</a>	05-28-2013 1:37 PM	05-31-2013 1:37 PM	No	No	

Screen 8:  
Administrator  
view of  
downloads

DataTrans Terms of Service Contact Us NCAASE geraldtd

**NCAASE** National Center on Assessment and Accountability for Special Education  
*Advancing research on growth measures, models, and policies for improved practice*

Users Uploads Downloads Expired

File Name	File Size	Date Downloaded	IP Address
-----------	-----------	-----------------	------------

## Screen 9: Administrator view of expirations

DataTrans									
<a href="#">Terms of Service</a> <a href="#">Contact Us</a> <a href="#">NCAASE</a> <span>geraldt</span>									
<b>NCAASE</b> National Center on Assessment and Accountability for Special Education <i>Advancing research on growth measures, models, and policies for improved practice</i>									
<a href="#">Users</a> <a href="#">Uploads</a> <a href="#">Downloads</a> <a href="#">Expired</a>									
Uploaded By	Name	Size	Link	Date Uploaded	Date Expired	Date Archived	Was Nonce?	Had Password?	Recipients
Trevor Cords	Less.zip	3.08 MB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-27-2013 12:45 PM	05-24-2013 12:46 PM	Yes	Yes	tcords@uoregon.edu
Gerald Tindal	DistricteasyCBMEnhancementsV2 copy.pdf	1.03 MB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-23-2013 9:43 PM	05-24-2013 12:40 PM	Yes	Yes	aaron.glasgow@me.com geraldt@uoregon.edu
Trevor Cords	ajax-loader.gif	4.67 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-23-2013 10:41 AM	05-24-2013 12:40 PM	No	No	aaron.glasgow@gmail.com tcords@uoregon.edu
Gerald Tindal	DistricteasyCBMEnhancementsV2 copy (1).pdf	1.03 MB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-23-2013 9:43 PM	05-22-2013 9:45 PM	Yes	Yes	aaron.glasgow@me.com geraldt@uoregon.edu
Trevor Cords	timesheet_UO.doc	60.5 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-22-2013 4:21 PM	05-22-2013 9:44 PM	Yes	Yes	
Aaron Glasgow	100mb.test	100 MB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 11:10 AM	05-22-2013 9:44 PM	No	No	
Aaron Glasgow	Screen Shot 2013-05-16 at 7.38.50 AM.png	172.84 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 11:14 AM	05-22-2013 9:44 PM	Yes	No	
Trevor Cords	timesheet_UO.doc	60.5 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 2:14 PM	05-22-2013 9:44 PM	No	Yes	
Trevor Cords	brt&ncaase_filetrans_system_updates_techspecs (1).docx	120.08 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 7:33 PM	05-22-2013 9:44 PM	Yes	No	
Trevor Cords	brt&ncaase_filetrans_system_updates_techspecs (2).docx	120.08 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-21-2013 7:36 PM	05-22-2013 9:44 PM	No	No	
Trevor Cords	brt&ncaase_filetrans_system_updates_techspecs (1).docx	120.08 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-21-2013 7:39 PM	05-22-2013 9:44 PM	No	No	
Trevor Cords	timesheet_UO.doc	60.5 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 9:11 AM	05-20-2013 10:41 AM	No	No	
Aaron Glasgow	Screen Shot 2013-05-16 at 7.38.11 AM.png	146.85 KB	<a href="#">Download</a>	05-28-2013 1:53 PM	05-20-2013 10:10 AM	05-20-2013 10:41 AM	No	No	

## CHAPTER 3

### DICTIONARIES FOR STATE AND EASYCBM DATASETS

The intent of this chapter is to be a dynamic repository for descriptions of data from all measures in all states that is updated as new data are acquired from state agencies and other sources. For some files, the variables may change from year to year, while others remain consistent over time. Documentation follows for the variables created by researchers for analyses as well as those that are part of the original data files, regardless of source.

The North Carolina datasets include data from several NCERDC and NC DPI files, along with additional variables created specifically to inform analyses. For years that contained anomalies that could not be easily resolved, variables were added from the exceptional children's services files, DPI's demo dataset, and the DPI test score dataset as an additional source to use as a crosscheck for duplicate cases and inconsistencies. All variables were retained from the original North Carolina test score files, with appropriate suffixes on all variables to distinguish between years (for example, the variable readscal for 2003 was labeled readscal\_03).

As part of processing these data, there was a search for duplicate records and an attempt to resolve them. Individual records with problems were tagged using year-identifiers so that a search for records that presented problems across multiple years could be completed. These decision rules are clarified in the descriptions of variables duptag\_XX, duptag\_mbXX, dupresolution\_XX, and doubledup\_XX.

*<See Appendix 3.1>*

Data obtained from the Oregon Department of Education (ODE) consist of student scores on both the Oregon State Assessment and the Oregon Alternate Assessment for school years

2004-2005 through 2010-2011. The original general education test contains all years in one file (with total scaled score for the alternate assessment), while the original alternate assessment test is separate for each year and also contains raw scores for items and tasks as well as scaled scores. For both the general education test and the alternate assessment, a separate file for 2011-2012 is available for integration into both original datasets. This file was separated by test type for some analyses, depending on the research questions addressed. Variables included in master dataset in its original form are described in the State Code Book for Oregon.

*<See Appendix 3.2>*

The Oregon Alternate Assessment was organized into files by year in order to combine additional information about students, such as item-level responses and task totals, with information in the state master file. A master file was created for each year of testing from school year 2007-2008 to 2011-2012. The test structure in those years was similar enough to make valid comparisons of a student's scores across years. From this yearly file, data necessary for longitudinal analyses were merged together to create a file of student scores from 2007-2008 to 2011-2012. Variables included in each year's master file are discussed in Appendix 3.3.

The easyCBM<sup>®</sup> datasets organize data in different ways for different purposes. The Benchmarks Item-Level (By Grade) dataset is a wide file in which each student is represented by one case. It includes student demographic information, benchmark scores for all easyCBM<sup>®</sup> measures for fall, winter, and spring, as well as item-level information for the Multiple-Choice Reading Comprehension, Vocabulary, and Math benchmarks. The item-level information not only includes whether or not students correctly answered questions on those measures, but also the options selected by students.

*<See Appendix 3.4a>*

The easyCBM® All Measures in System file is a long file in which each case represents a measure taken by a student. Variables in this file include student demographic information, whether the measure was a benchmark assessment or used for progress monitoring, the type of measure, and the date taken.

*<See Appendix 3.4b>*

The easyCBM® Interventions file is a long file in which each case represents a measure taken by a student. Variables in this file include student demographic information, subject in which the student is receiving an intervention, type of intervention, and, in some cases, method in which the intervention was administered.

*<See Appendix 3.4c>*

The District Names from Riverside file illustrates all of the different identifying information for districts using easyCBM® with students. This file includes a unique identifier for each district the state abbreviation, and the district name. Schools in Oregon also have a 4-digit state/district ID number included.

*<See Appendix 3.4d>*

## CHAPTER 4

### PROCESSING DATA SETS

Analysis by multiple researchers using large data sets from several states requires that the data are cleaned and structured using clear processes. This ensures both easier accessibility of the data to all researchers involved in the project and also increased opportunities to replicate types of analyses conducted on one state’s data in another state. Procedures for cleaning and arranging data, while different from dataset to dataset due to their characteristics and original formatting, were modeled after conventions for making cross-state comparisons developed by Goldschmidt, Choi, and Beaudoin (2012)<sup>1</sup>.

<See Appendix 4.1>

While the intent of this project is not to make cross-state comparisons, efficiency increases when replicating analyses done using one state’s data in another state if the conventions for cleaning data are well documented and follows clear rules. This facilitates easier applying methods used in one state to the data from another state, as the procedures used to clean and process both datasets are easily interpretable. We summarize procedures used for processing data for each specific data set separately, to highlight each dataset’s unique characteristics. We also document as the specific procedures used with each in a series of appendices to ensure maximum usefulness in later analyses.

*North Carolina.* Building the North Carolina files<sup>2</sup> involved steps to vet the accuracy of data in the files, as well as to ensure the appropriateness of data for use in analyses. To determine

---

<sup>1</sup> Goldschmidt, P., Choi, K., & Beaudoin, J.P. (2012). *Growth Model Comparison Study: Practical Implications of Alternative Models for Evaluating School Performance*. Washington, DC: The Council of Chief State School Officers.

<sup>2</sup> Procedure created and conducted by Dr. Schulte, North Carolina State University.

data in each year's raw data files were accurate, they were compared to the appropriate year's North Carolina *Green Book*<sup>3</sup>, an annual state publication of testing results available online from school year 1997-1998 to 2010-2011. Subsequent elimination of duplicate records took place not only within files, but also across grades, which necessitated extra attention to the records between different grade-level files within the same year to avoid inaccurate data. Coding variables in each year's dataset by year helped to ensure that student records were properly matched for longitudinal analyses.

<See Appendix 4.2a and 4.2b>

Before deleting any duplicate records, they were compared. Discrepant variables (e.g., one record says child got extended time on a test, one doesn't) were changed to err on side of inclusiveness for EC services, accommodations, free lunch, and LEP.

Further cleaning of the North Carolina data was conducted<sup>4</sup> to create a longitudinal cohort beginning with year 2001. Data were inspected to identify cases missing information central to planned analyses (e.g. sex, ethnicity) and appropriate cases were dropped from the files. See appendix for specific description of cases eliminated, as well as numbers of cases this impacted.

< See Appendix 4.3>

*Oregon Alternate Assessment.* Although data were available for the Oregon Alternate Assessment files for school year 2004-2005 through 2011-2012, only data from years 2007-2008 through 2011-2012 were appropriate for combination into a longitudinal dataset. Previous years included significant changes to the structure of the measures, rendering them

---

<sup>3</sup> Access the Green Book for different years from the following website:

<http://www.dpi.state.nc.us/accountability/testing/reports/archive>

<sup>4</sup> These steps in data cleaning were created and conducted by Dr. Stevens, University of Oregon.

meaningless for comparison of individual students' scores across years. The structure of each raw data file was different, necessitating standardization of variable names and file layout before constructing longitudinal datasets of student scores<sup>5</sup>. For test years lacking specific disability codes for students, those values were merged in from a master list of students with disabilities provided by the Oregon Department of Education (ODE) for each school year.

*<See Appendix 4.4>*

The final files for each testing year from 2007-2008 to 2011-2012 represent a combination of data from the raw data file and data file from ODE that include demographic variables not available in the original raw file such as indicators of limited English proficiency (LEP) and the final scores reported for AYP purposes for applicable students. A full list of all variables in the alternate assessment files is included in Appendix 4.4.

*easyCBM Reading and Mathematics.* The easyCBM data<sup>6</sup> presented a set of challenges distinct from the state datasets; the student demographic information present in the easyCBM data files was voluntarily inputted by a variety of individuals (depending on the school or district) rather than being formally inputted at the district or state level. To add accurate demographic information to the easyCBM measures in grades 3-7 for as many students in Oregon as possible, student demographic data were collected from Oregon school districts and subsequently added to the easyCBM files. In this fashion it was possible to obtain accurate demographic information about students in Oregon who took these assessments, rather than relying on demographic information in the original file that may, in some cases, have been entered by an individual unfamiliar with the student's formal records. Steps were taken before

---

<sup>5</sup> These steps in data cleaning were created by Dr. Tindal and conducted by Jessica Saven.

<sup>6</sup> Steps in data rendering created by Dr. Tindal and conducted with assistance of Jessica Saven.

the addition of this demographic information to remove duplicate cases and to eliminate variables that were not relevant to any future analyses. Benchmark and progress monitoring files were combined into one dataset in order to allow a more comprehensive analysis of within-year student growth.

*<See Appendix 4.5>*

For a subset of easyCBM files (Passage Reading Fluency<sup>7</sup>) used for progress monitoring, data cleaning used a set of process rules specific to this dataset. First, all out of range scores were deleted (e.g., < 0 and > 400). A rule for expected growth was applied to identify invalid scores for deletion, which is summarized as follows: if two scores for one student were less than two weeks (14 days) apart **and** different by more than 35 words correct per minute, the score least like adjacent scores was deleted. Additionally, if scores in the dataset were within 7 days of each other, the median of the scores was used. If ORF scores were more than 7 days, but less than 28 days apart, both scores were kept. Scores more than 28 days apart were deleted.

Students were deleted from grades other than 1-8, and also if they had any instance of off-grade-level testing, defined as taking a test one or more grades above or below their own grade level. Additionally, students were deleted who scored above the 50<sup>th</sup> percentile on the benchmark or first progress monitoring occasion in fall, winter, or spring. Finally, students with less than three progress monitoring tests were deleted.

*<See Appendix 4.6>*

---

<sup>7</sup> Decision rules created and applied by Dr. Nese, University of Oregon.

## **CHAPTER 5 STRUCTURING DATA SETS**

A number of issues need to be considered when structuring data for analysis.

(1) Types of State databases to be constructed. Our commission is to focus on students in grades 3-8 by year, all kids, all types of assessments. However, do we include kids not present at end of year (am working to do that in NC)? Our next commission is to create longitudinal data sets by first placement in 3rd grade through 8th grade (using a marker for all students, including alternate). However, do we use only kids who were present in 3rd grade? Finally, if we are to include school outcomes, should they be created from datasets or state data or both?

(2) Uniform criteria for accuracy are needed. NCAASE could use standards set for NC

(3) Naming conventions are needed. Do we keep original data names to be able to use state codebooks? Do we have state codebooks? Do we have universal variable names or a crosswalk?

(4) Data need to be verified: Do we verify our data against state figures?

(5) Documentation needs to be created: How do we document who was included in our databases? Can state departments provide feedback who was left out? Do we have uniform decisions?

### Processes to Use in Creating Data Sets

A web application can be used from a desktop for accessing the databases OR the database may be accessed directly; both can be used interactively to complete essential data queries and builds. The server setup for our training use is: [dbintro.com](http://dbintro.com)

This site supports ssh connections and the best clients are either the UofO provided tools (PCs) or terminal and fetch (Mac OS). Users need an account (established by Aaron Glasgow); all accounts require a user name, based on uoid (the part before your email institution reference) and a password. This login is based on the uoid however is not associated with it. The login for ncaase SQL data servers will be uniquely assigned and for clarity will be referenced as ncaaseid in this document. This server is available 24/7/365.

### Accessing Files through a Terminal and Command Lines

Through a terminal program (for a MAC, this is available in Applications...Utilities...Terminal) the interaction begins by logging in over a secure shell (prompts are italicized and responses are underlined). For PCs: there is a free download at <https://it.uoregon.edu/ssh-secure-shell>. Interactive development of files from the terminal is a good to inspect data sets; the command lines can then be used to actually build files and export them.

**Step 1.** Connect to the site using the terminal

```
$ ssh dbintro.com -l ncaaseid  
ncaaseid@dbintro.com's password:
```

*Linux bvm1 2.6.32-5-amd64 #1 SMP Sun Sep 23 10:07:46 UTC 2012 x86\_64*

*Last login: Wed Nov 7 08:09:41 2012 from 63-155-154-201.eugn.qwest.net*

Note: Examples from this point forward reference the ncaaseid gerald, substitute your unique ncaaseid in actual use.

*geraldt@bvm1:~\$*

**List Files in current directory**

```
geraldt@bvm1:~$ ls
jtstudents.txt  students.csv
```

**Step 2.** Begin mysql monitor session

```
geraldt@bvm1:~$ mysql -u gerald -p
Enter password: _____
```

*Welcome to the MySQL monitor. Commands end with ; or \g.*

*Your MySQL connection id is 114*

*Server version: 5.1.63-0+squeeze1 (Debian)*

*Copyright (c) 2000, 2011, Oracle and/or its affiliates. All rights reserved.*

*Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.*

*Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.*

**Step 3. Show available databases**

```
mysql> show databases;
+-----+
| Database      |
+-----+
| information_schema |
| geraldtdt      |
+-----+
2 rows in set (0.00 sec)
```

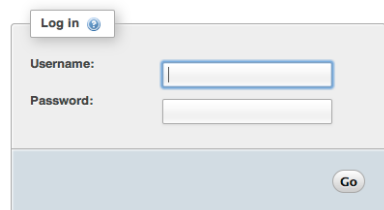
**Step 4. Access a database**

```
mysql> use geraldtdt;
```

At this point, various commands can be used to view or manipulate data contained in various tables.

**Step 5: Exit from mysql monitor to shell**

```
mysql> exit;
```

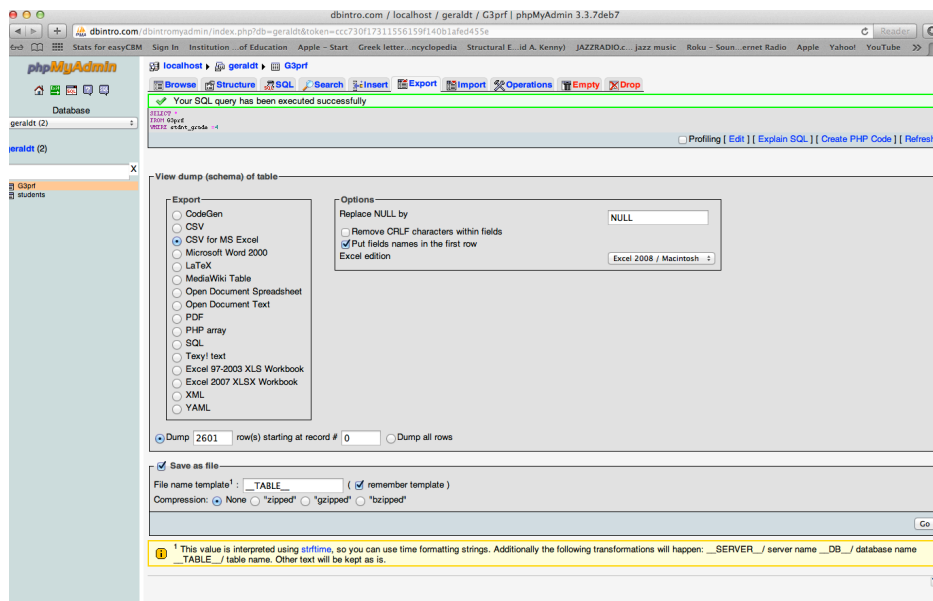
**Accessing Files through phpMyAdmin**

<http://dbintro.com/dbintromyadmin/> (a temporary server)

Open a web browser and navigate to the URL. Use your ncaaseid and database password to gain access to the phpMyAdmin database management software. Interactively create commands in the terminal to see how they manipulate the data; once satisfied, the final set of commands may be copied and pasted into the open screen in the actual sql database for eventual export.



## Export file in CSV format



## Links for Further Training Support

Official MySQL Reference <http://dev.mysql.com/doc/refman/5.1/en/index.html>

Google Code University "Introduction to Databases and MySQL" - Another Nice Overview

<http://code.google.com/edu/tools101/mysql.html>

A "cheat sheet" of common SQL commands. These have been released to the public domain

<http://en.wikibooks.org/wiki/MySQL/CheatSheet>