BSD Certification: How to Create a Psychometrically Valid Certification Examination

Dru Lavigne, EuroBSDCon 2011

The BSD Certification Group Inc. (BSDCG), founded in 2005, is a non-profit organization committed to creating and maintaining a global certification standard for system administration on BSD based operating systems. One of the founding tenets is that its certifications are psychometrically valid in order to provide value to both the system administrator and the employer. This paper provides a brief introduction to psychometrics and its value proposition. The remainder of this paper outlines the process of creating and maintaining a psychometrically valid examination and how members of the BSD community can contribute to this process.

What is Psychometrics and Why Should I Care?

The primary goal of any certification organization is to accurately assess a specified set of knowledge and skills. But how is that set of knowledge and skills defined? And how can the accuracy of the assessment be determined? Through the science of psychometrics (http://en.wikipedia.org/wiki/Psychometrics). Psychometrics isn't something that just occurs once; it is an ongoing process that begins in the planning stages of the certification and continues after the examination becomes available.

The goal of a psychometrically valid assessment is NOT to determine if a candidate knows the answers to particular items, but rather if the examinee can demonstrate understanding of the concepts behind the items. This is the reason why the BSDCG does not provide sample exam questions: if an examinee understands a concept, it does not matter what question is asked, as long as the question is clearly stated and matches the concept being assessed.

Once you know what to look for, it is fairly easy to determine if an examination is psychometrically valid. Some of the features of a pyschometrically valid assessment are as follows:

1. **Published examination requirements:** the learning objectives to be tested are publicly available and clearly define the specific tasks to be measured by the assessment. The skill level (audience) that is required to pass the assessment is clearly defined. Learning objectives are matched to knowledge domains and the percentage (importance) of each domain is identified, with all percentages adding up to 100%. The BSDCG has published requirements for two certifications: the BSD Administrator (BSDA, http://www.bsdcertification.org/downloads/pr_20051005_certreq_bsda_en_en.pdf) and the BSD Professional (BSDP, http://www.bsdcertification.org/downloads/certreq_bsdp_en.pdf)
2. **Examination matches the published requirements:** the assessment does not contain any items that do not map to an exam objective or which are too easy or too hard for the defined audience. The number of items for each knowledge domain reflects the domain's percentage: for example, an examination can not contain 60% of questions covering a knowledge domain that is worth 15% of the examination.
3. **The assessment is repeatable and defendable:** for example, if an examinee

was to take 2 different versions of the assessment on the same day, that person should receive similar results. Similarly, an examinee who meets the intended audience requirement and who understands the exam objectives should not receive a lower score than an examinee who does not meet the minimum skill requirements and who does not understand the exam objectives. Psychometrics uses statistical analysis to determine trends and find anomalies by examining how every examinee responds to each examination question.

4. **Does not introduce bias:** for example, to examinees whose native language is not English. Exam questions need to be worded as clearly as possible as the goal of the assessment is to determine if the examinee understands the concept, not to determine if the examinee can figure out what question is being asked.
5. **Does not try to trick the examinee:** for example: by asking questions not covered by the exam objectives, by using ambiguous wording or providing distracting details which are irrelevant to the objective, or by asking the examinee to select multiple correct answers from many (e.g. offer 11 possibilities and ask the examinee to select "all that are correct").
6. **Does not contain inaccuracies:** such as multiple possibly correct answers that vary by interpretation, questions or answers that are technically incorrect, and formatting or grammatical errors.

These features of a psychometrically valid examination increase the value of the assessment for the examinee seeking certification, the employer who hires a certificant, the organization that maintains the certification, and the community that deals with the intended audience of the certification. The value to each can be summarized as follows:

Certificant: an examinee who seeks to be certified is assured that the assessment will only ask questions that cover the examination's objectives, that the questions asked will not be at a skill level higher than the intended audience, and that the number of items to be asked in each knowledge domain will be reflective of that domain's percentage. This means that even if additional study aides are not available, examinees can still gauge their exam preparedness by comparing their skillset to the listed objectives. This also means that the examination requirements document is the definitive study aide.

Employer: an employer who hires certificants or who encourages existing employees to take a psychometrically valid examination has a yardstick that validates the minimum skillset of certified employees. Any employee who passes the examination has demonstrated that they understand the skills listed in that examination's objectives. For this reason, employers are an integral component of the examination creation process as they can provide input regarding which tasks are important to an employer.

Certifying Organization: an organization that provides pyschometrically valid examinations is in a position to successfully defend the reliability of the assessment should an examinee dispute the results of their examination. Additionally, internationally recognized accreditation standards, such as the ANSI/ISO/IEC 17024 Accreditation Program for Personnel Certification Bodies (https://www.ansica.org/wwwversion2/outside/PERgeneral.asp?menuID=2), include psychometrics as part of the requirements for accreditation.

Community: the community as a whole benefits when exams are psychometrically valid as the certificate is worth more than the paper it is written on and certificants are not able to pass the exam by memorizing the content available from "exam cram" websites. This increases the value of the assessment and the respect given to certificants.

The rest of this paper describes the steps in the process of creating a psychometrically

valid examination.

The Job Task Analysis (JTA)

The first step in the exam creation process is to determine which tasks should be assessed. This requires a survey of those who will benefit from the certification: the intended audience (those who will take the exam) and the employers who hire them. This survey is known as a Job Task Analysis (JTA). Additional surveys may also be required to determine the community's needs: which certifications are desirable, the skill level to be assessed by each certification, and which delivery method is most appropriate for the certification. For example, the BSDCG's preliminary survey of the BSD system administrator community determined that two certification programs were needed: one that assessed a junior level administrator with a minimum of 6 months experience (the BSDA examination) and one that assessed a senior level administrator with a minimum of 2-3 years experience (the BSDP examination). Subsequent surveys helped to fine-tune the needs of the BSD system administrator community. The Reports from these surveys are available from:

- http://www.bsdcertification.org/downloads/sr1_links.pdf
- http://www.bsdcertification.org/downloads/delivery_survey.pdf
- http://www.bsdcertification.org/downloads/pr_20051031_usage_survey_en_en.pdf

The goal of the JTA is to identify the content areas (domains) and the importance and frequency of each task the intended audience is expected to know. Ideally the sample size of the JTA is large (several hundred or thousand people should take the survey) and targeted towards both the intended audience and their employers.

An example of several surveyed items from the BSDP JTA Report is shown in Figure 1 of the Appendix.

Certification Requirements

Once the results of the JTA have been gathered, they are analyzed to determine which tasks are considered important and frequent enough to be included on the examination. This analysis results in the Certification Requirements document which should be published and freely available to examinees, employers, and the creators of study aides. This document contains the following information:

1. **A clear definition of the intended audience.** For example, the audience for the BSDP is defined as: *The BSDP candidate is experienced in using one or more versions of NetBSD, FreeBSD, OpenBSD, or DragonFly BSD for several years. The candidate may be engaged as a senior BSD system administrator by an employer, may be active in open source projects where BSD systems are used, or may have gained experience in using BSD systems over the years in a variety of settings. The candidate generally has a well developed knowledge of shell scripting, networking, user administration, file and filesystem maintenance, kernel tuning, and many other technical tasks.*
2. **The knowledge domains and their percentages.** These are typically listed as a Table of Specifications. The Table of Specifications for the BSDA examination is shown in Table 1 of the Appendix. It clearly defines 7 knowledge domains and the percentages for all domains add up to 100%. This table is sometimes called the Examination Blueprint as it defines the domains to be assessed and the

percentage of questions to appear for each domain.

3. **Detailed exam objectives**, where each objective clearly identifies the task or concept that will be assessed. To continue with the example from Figure 1, Figure 2 in the Appendix shows the exam objective for the first item on the JTA: "Build and install system from source code".
4. **The depth of knowledge to be assessed according to Bloom's Taxonomy** (http://en.wikipedia.org/wiki/Blooms_taxonomy). As an example, most of the items in the BSDA exam are knowledge-based, while many items on the BSDP exam are comprehension- or application-based.

The examination objective shown in Figure 2 in the Appendix clearly states the required task, its importance and frequency as determined by the JTA, that both knowledge and comprehension of this task is expected of the examinee (Bloom: K,C), a description of the concept behind the task, which commands can be used to perform the task, and the maximum time limit should the question ask the person to perform the task in a lab-based scenario.

Item Writing

Once the Certification Requirements have been published, SMEs (subject matter experts) can use the Blueprint to begin to write items (questions) for the examination. An SME has experience in the concepts to be assessed by the examination and works under the terms of an NDA (non-disclosure agreement) to protect the confidentiality of the examination's items.

In the example of a multiple choice examination, every exam item contains the following elements:

- **Stem:** the question being asked.
- **Key:** the correct answer. It is listed as one of the options that follows the stem.
- **Distractors:** three or four incorrect answers that appear as options that follow the stem.

When writing examination items, SME's must keep the following points in mind:

- Items must be written to the skill level of the audience defined in the Certification Requirements.
- Items must be clearly written and understandable, even for examinees who are non-native English speakers. The examinee should not have to "guess" what the question is asking or wade through irrevelant information not related to the objective being assessed.
- Each item assesses only one objective. Questions that assess multiple objectives should be rewritten as separate items.
- Items should be consistent in formatting style, punctuation, and grammar to avoid distracting the examinee.
- Avoid items that require the examinee to select multiple answers or that include an answer of "all of the above" or "none of the above".
- If instructional materials exist, do not cut/paste information to create items. This offers an unfair advantage to those who can afford the materials and encourages memorization rather than understanding.

- Do not place enemies on the same exam form. Enemies are items that contain information that can be used to answer another question on the exam.
- Distractors should be of similar length and plausibility as the correct answer. Someone who does not understand the objective should not be able to pick out the correct answer by using contextual cues: for example, the correct answer is of a different length, the distractors are completely inplausible, or the correct answer contains words found in the stem.
- Avoid stems containing negatives or double negatives as these can be confusing, especially to non-native English speakers.
- Have a standard for determining the order of the options (e.g. commands are alphabetized). Ensure that the correct answer is not always in the same position; for example, that answer C is not always/usually the correct answer. Avoid having the first and last option as always being incorrect.

It should be noted that the first draft of an item rarely looks like the final version of the question that appears on the examination. Once a pool of items that address most of the examination's objectives have been written, a lengthy review process begins. Every item is reviewed by multiple SMEs and the psychometrician to ensure that the question is clear, unambiguous, and does not contain any clues to assist the examinee in guessing the correct answer. In practice, most items undergo significant rewrites during the review process.

Once the questions are as clear as possible, the psychometrician assists in creating at least one form (version) of the examination. Most organizations create multiple forms to prevent examinees from memorizing exam questions. The psychometrician ensures that each form, even though it contains different questions, results in an equal assessment so that it does not matter which version of the exam an examinee takes. This requires an analysis of item difficulty and ensuring that the domain percentages defined by the Table of Specifications are adhered to. Once the forms are created, the beta period can begin.

Beta Period

The goal of the beta period is to generate statistical information on how actual examinees respond to the exam questions. Ideally, the sample size of the beta exam is at least 100 people (the more, the better), is comprised of people who match the definition of the intended audience, and includes beta examinees from each country where the certification will be offered. The purpose of the beta period is to identify any anomalies in how beta participants respond to the exam's questions.

For example, if 95% of the examinees answer a particular question correctly (or incorrectly), it should be reviewed to see if the item is too easy (or hard) for the audience or if it contains contextual cues pointing to the correct answer (or one of the distractors).

As another example, if a particular item is answered differently by English speakers than non-native English speakers, the wording of the item may not be clear or may contain a colloquialism that only makes sense to native English speakers. The item needs to be fixed in order to prevent unfair bias towards non-native speakers.

It should be noted that a beta participant can not "pass" or "fail" the beta exam as the passing score can not be determined before the Angoff session. Instead, the statistics gathered by the beta exam are used to determine if any items need to be further

reviewed and rewritten before they can be considered for inclusion in the final version of the examination. The certifying organization needs to consider the ratio of beta examinees to potential exam candidates and how many questions are revised between the beta exam and "real" exam to determine which processes are necessary to protect the confidentiality of the exam (e.g. using multiple exam forms, not allowing beta examinees to take the "real" exam, or ensuring that a beta examinee receives a different form should they take the "real" exam).

Angoff Session

Once the statistics from the beta period have been analyzed and the necessary exam items have been reviewed and rewritten, the psychometrician and SMEs will perform an Angoff session (http://en.wikipedia.org/wiki/Standard-setting_study). The goal of the Angoff session is to set the cut score (passing score) by determining the number of items a minimally competent examinee would answer correctly. Anyone who correctly answers more items passes the exam as they will have demonstrated that they are more than minimally competent.

During the Angoff session, each participant reviews every exam item and rates it on a scale of 1-10 on the likelihood that a minimally competent person would answer that question correctly. Once all the items are individually rated, the ratings are compared and any discrepancies are discussed. For example, if one SME thinks the likelihood of getting the answer correct is 2 and another SME thinks the likelihood is 8, each has to explain their reasoning. After a period of discussion, the group as a whole reaches concensus on one rating per item. The psychometrician uses the list of ratings to determine the passing score for the exam. Once the passing score is determined, the "real" examination is ready to be published so that it can be taken by examinees.

Exam Maintenance

Once an examination is published, the psychometric process does not end. Statistics continue to be gathered for analysis on how examinees respond to each exam item. The certifying organization will work with the psychometrician to determine how often it is appropriate to review the statistics for an examination. Typically, a review will occur after so many attendees take the exam (e.g. every 100, 500, or 1000 examinees) or at a set period of time (e.g. every 6 or 12 months). The review period should take into account how many examinees take the exam as there needs to be sufficient examinees to gather useful statistics. Additionally, the organization does not want to let long periods of time lapse between reviews, especially if many examinees take the exam within a time period.

For each review period, the psychometrician will identify which items should be reviewed by the SMEs. These will include items that are disproportionately answered correctly or incorrectly by most examinees as well as any items that are statistically strange (e.g. items which minimally competent examinees tend to answer correctly while very competent examinees tend to answer incorrectly). The SMEs may choose to rewrite these items or replace them with different items. The psychometrician may also decide to introduce "pilot" items which don't affect the examinee's score. The purpose of the pilot items is to determine how examinees respond, without having to go through another beta period and Angoff session. Once the review is complete, the certifying organization will publish the new version of the exam. Many organizations do not publicize that the exam has changed in order to protect the confidentiality of the items within the exam. It is important to remember that, due to the psychometric process, the

new version of the exam is defensible and test-independent in that if the same examinee was to take both the old version and new version of the exam, they would receive similar scores.

The statistical analysis can also be used to spot anomalies that indicate that the confidentiality of the exam has been compromised. For example, if the exam scores for a particular testing center or exam proctor are consistently higher than the statistical average, the certifying organization needs to launch an investigation to determine the reason why this trend is occurring.

Becoming Involved

There are many ways that members of the BSD community can contribute to the certification effort. Here are some suggestions for those interested in becoming involved:

BSD system administrators: BSD certification exams assess real-world, on the job skills and the input of working administrators helps to ensure both the quality and reliability of the exam items. System administrators can act as SMEs and participate in item writing, item review, and Angoff sessions. System administrators are also encouraged to take the exam and to indicate that they are BSD certified. Even if you don't need the certificate to gain employment, your participation helps to grow the number of certificants and to increase the awareness of BSD certification to employers.

Employers of BSD system administrators: including BSD certification in job postings helps to increase awareness of the certification program and assists your organization in finding suitable candidates for system administration positions.

BSD advocates: anyone who is interested in promoting BSD can assist the BSDCG in spreading the word about BSD certification. Does your school or employer know that a certification program exists? What about your local user group? Can you blog about BSD certification, make it known in certification forums, write an article for a magazine or ezine, or give a presentation at a conference or local event?

You can learn more about BSD certification, download our publications, subscribe to the mailing list or social media sites, and get more information on how to take the exam from the BSDCG website: http://www.bsdcertification.org.

Summary

Creating and maintaining a quality certification program is a lengthy and time-consuming process. The science of psychometrics can guide a certifying organization through this process by providing a series of distinct and repeatable steps. Psychometrics helps to maintain the quality of the examination's questions, allowing for a reliable assessment of the skills being certified. This provides great value to those seeking certification, the employers who hire certificants, and the organization that provides the certification.

Appendix

Figure 1: Sample of Response to Items in JTA

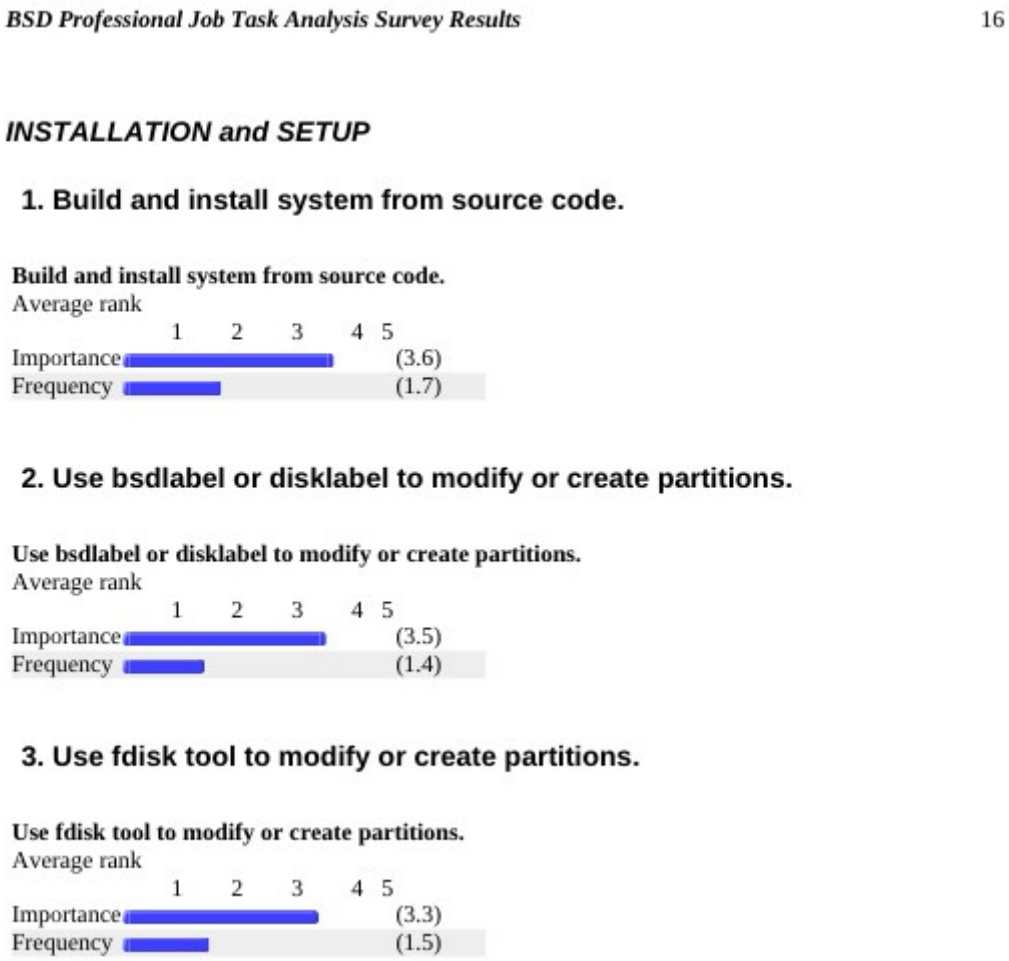

Table 1: Table of Specifications for BSDA Examination

Domain	Weighting
Installing & Upgrading the OS and Software	13%
Securing the Operating System	11%
Files, Filesystems, and Disks	15%
Users and Accounts Management	16%
Basic System Administration	12%
Network Administration	15%
Basic Unix Skills	17%

Figure 2: An Example Objective in a Certification Requirements Document

3.1 Installation and Setup

System installation and setup is a basic but important job of a BSD system administrator. Beyond getting the system up and running, the BSDP candidate must be able to perform more complex tasks such as preparing and updating automated installation scripts, using a serial console for installation, restoring a corrupted boot sector, and other similar tasks.

3.1.1 Build and install system from source code.

Importance: 3.6, Frequency: 1.7, Bloom: K,C

Concept:

BSD systems allow the administrator to compile both the kernel and userland from source in order to upgrade an existing operating system. The commands provided for this purpose vary between the BSDs.

BSDP candidates should be familiar with the steps required to compile both kernel and userland from source and to install the new kernel and userland for their BSD of choice.

Practical:

DragonFly BSD: build(7), development(7), make(1)

FreeBSD: config(8), make(1), mergemaster(8)

NetBSD: build.sh, etcupdate(8), postinstall(8)

OpenBSD: config(8), make(1), sysmerge(8)

Estimated Time:

10 minutes to describe the steps required to build and install a system from source code.