



Beer Judge Certification Program

5115 Excelsior Blvd, #326
St. Louis Park, MN 55416

www.bjcp.org

Guidelines for Lead Graders of BJCP Exams

Revised: February 2008

Introduction

As a lead grader of BJCP exams, you are responsible for providing feedback to examinees on their performance on the qualification exam for our program. This is often the first, and quite possibly the only, such feedback that many new judges will ever see. It is therefore vitally important that the critique be carried out in a positive manner with no cynical, derisive or sarcastic comments. We want to support our new judges and encourage those who fail to try again. This is often a difficult task, especially with examinees who were unprepared.

Lead graders have a secondary responsibility to support the training and development of new graders. Since grading is somewhat of a “bootstrap” system where new graders learn the process from lead graders, the responsibility to help new graders develop and become more effective rests primarily with lead graders. This duty should not take precedence over completing the feedback in a timely manner, but it should also not be ignored.

Since lead grading plays such an important role in the Exam Program, several senior graders with substantial experience documented the best practices, lessons learned, and practical tips that they frequently share with new lead graders. These perspectives on grading are meant to answer common questions, provide insight based on experience, and generally provide additional educational and training assistance that new lead graders often desire. These guidelines provide practical guidance on how to approach the lead grader role, and describe what is expected of you in this critical job.

Scoring and Consensus

One of the main goals of exam grading is to arrive at fair and accurate exam scores. As soon as the lead and the second graders have completed grading the exam set, it's important to work toward a consensus score. Usually this begins with an exchange of raw scores, often by sending completed Exam Grading Forms to each other. A simple list of written, tasting and total scores can also be exchanged.

Many lead graders do a quick scan of the raw scores to see if there are any major disagreements between themselves and the second grader. The purpose of this review is to identify those exams that may need additional review by both graders. If the scoring on all exams is close, this may not be needed. If the scoring is uniformly distant, then the graders should discuss general grading philosophy and agree on a scoring approach. If only a few exams are apart in scoring, then those exams should receive additional attention from the graders.

In many cases, especially when the scores are in the same judge level range, a simple numeric average may be used as the consensus score. However, this is not a requirement. Either grader may choose to move in the direction of the other grader, as is often done when judging beers in competition. If the averaging approach is used, note that it is important to average only the written and taste portions of the exam (not the individual questions) and then calculate the resulting overall score from these section consensus scores.

If a more serious disagreement in scores exists, it is worthwhile discussing general impressions or scores for particular exam questions (or beers, on the Tasting Section). The Exam Grading Form from each grader can be used to pinpoint scoring differences on individual questions, which can then be discussed in more detail. Often a discussion centering on what the question requires is helpful, particularly on the point allocation of each required component of a question. Different grading techniques may be used by individual graders, but the final scores should still reflect the knowledge level demonstrated on each question.

If an agreement cannot be reached, or if there are remaining doubts, ask the Associate Director to closely review or grade troublesome exams. Do not let a disagreement on a small number of exams keep you from completing your task. Associate Directors often spot-check the scoring on exams anyway, so it is no problem to direct them to specific exams that need a second opinion.

Note that considerable leeway can be given when determining the score for the Scoring Accuracy component of the Tasting Section of the exam, as this is the most subjective part of the entire exam. Factors such as the skill of the proctors, the condition of the exam beers, and even serving conditions may affect the relative scoring. If the proctors' scoresheets are poor, or are in basic disagreement with the majority of the examinees, it is acceptable to override the automatic score generated by the Grader's Scoresheet by adding or subtracting a few points by modifying the assigned consensus score in the EGF spreadsheet.

Consider the skill level of the proctors. Examination guidelines require two judges capable of writing National or higher level scoresheets; however, Certified judges are often used. If an examinee is retaking the tasting portion of the exam, it is quite possible that this judge will be higher ranked than the proctors. Exam administrators should avoid this situation, but it is unfortunately rather common. Allow for the possibility.

It is also a good idea to calculate the average of the examinees' scores to compare it with the proctors' consensus score. Check the notes that were submitted by the Exam Administrator, consider the Administrator's impression of the beer, and note whether the Administrator doctored it. If a beer was doctored and detected by examinees but not proctors, don't be too hard on the examinees.

If three proctors are used, it is also acceptable to discount one proctor's score if it does not agree with the other proctors (and/or examinees). Look at each proctor's score and see how the proctor consensus was determined. If, in your judgment, you wish to discount the score of one of more proctors, calculate or estimate new consensus proctor scores to be used for grading purposes. These new scores can be input into the grader's spreadsheet to calculate new Scoring Accuracy components for the examinees.

It is a good idea to avoid assigning consensus scores that end in a '9,' especially the overall score. These will usually be adjusted by the Associate Director or Exam Director anyway, since these scores are more likely to result in a protested exam. If a score is borderline, whether for one section or the entire exam, review it closely to ensure that it is at the proper level. Verify that the exam is scored at the proper level, looking at it as a whole. Individual graders can have scores ending in '9,' since those are often averaged or adjusted. This recommendation applies to the final team consensus scores.

Scoring Master-Level Exams

If you look at the statistics for Master-level scores since the inception of the BJCP, you will find two striking points. The first is the astonishingly low number of examinees (roughly 2%) who have scored at the Master level. The second, and most important, is the fact that the average of all these exams is 90.7—very nearly the *minimum* score required for Master level. Lead graders need to understand these facts.

The ‘nature of the beast’ of open-ended exams is that they create an atmosphere where the examinee will write as much as possible on each given topic. With all of this content, there will inevitably be minor mistakes. Rather than going through specifically counting all the mistakes where you could deduct points, it’s helpful for lead graders take a step back and look at the exam as a whole.

It’s extremely important for a grader to understand that, from time to time, an examinee will have more experience and knowledge than a grader. These exams are usually written by someone going for a Master-level score and often have voluminous content. Don’t create artificial reasons not to give a Master-level score. Don’t find a reason not to give a score higher than a 90. True Master-level exams are a real joy to grade and also are frequently instructive for the grader.

Feedback: Content

It is important to be accurate when evaluating and critiquing an exam. If you have any doubt as to the correctness of a response, take the time to research it. Be aware that there is often not unanimous agreement between sources, so better answers will often acknowledge more than one point of view. For style information, the main resource should always be the latest version of the BJCP Style Guidelines, since these are the main study reference for most examinees. The exam wording has been reworked to refer specifically to this source.

Each examinee and grader brings different levels of knowledge, education and experience. A lead grader needs to understand, appreciate, and value this diversity. There will be times when you will get answers to questions that you may have never before seen nor heard. There may be times when an answer seems completely off-the-wall. In these instances, a lead grader should do a little research to see if these answers have merit. There are some highly advanced people in the BJCP that are on the cutting edge of brewing and brewing science. Just because an answer is complicated, doesn’t mean it shouldn’t be given its full merit. Ask the Associate Director for help if you don’t know whether an answer is correct, or are unable to find a reference to properly evaluate the response.

Feedback: Tone

Please be positive when providing feedback. You can do this in several different ways, but one effective technique is to emphasize what the examinee answered correctly before noting any miscues. This approach helps make the criticism easier to accept. As an example, rather than stating, “You missed three of 15 T/F questions,” try something like “You correctly answered 12 of 15 T/F questions.” In particular with the True-False questions, it is a good idea to recap the incorrect responses and point out the correct answers. Many weeks have elapsed between the when the exam was administered and when the feedback is received, so it’s safe to assume that the examinees will not remember accurately what they wrote.

It’s a good idea to begin the feedback with an overall statement of congratulations for those who passed, and encouragement for those who did not (alternatively, you might congratulate them on completing the exam, since this represents a significant effort). It helps to add a similar comment to the wrap-up statement of the Report to Participant (RTP), as well. Welcoming new judges to the BJCP and encouraging them to continue their training and development is also appropriate. A useful technique for

improving feedback is to liberally quote from the exam. Since exams are not returned, the inclusion of quotes can reduce the likelihood of a contested exam result.

Feedback: Format

It is imperative to provide complete comments to the examinee. The answer to each question should be assessed; one common way to do so is to use a list of bullet points, one for each of the Essay questions. Each exam beer should be addressed in a similar manner. Try to comment on the major aspects against which exam takers are graded: perception, descriptive ability, feedback, completeness and communication, and scoring accuracy. You are allowed to pass along the identity of exam beers in this feedback.

Note that assigned scores should correspond to comments made on the RTP, and that rank levels in the Essay Portion and Taste Portion boxes on page one of the RTP should be shaded to match the scoring and comments. This is similar to using the Scoring Guide on the BJCP scoresheet when judging beer. If an examinee aced a question, don't simply say that they did a "good job" – a perfect score is "outstanding" or a similar superlative. Likewise, you also wouldn't say that they did a "good job" if they totally botched a question. Match comments to scores, including the degree of praise or criticism used. Scores of individual questions should match the skill level demonstrated in the answer (e.g., a 9-10 score on a single question is a Master-level response; an 8-9 is a National-level response, etc.). The same score-to-rank mapping applies for the shaded areas on the first page of the RTP using the average of scores on technical questions, style questions, and the BJCP question, as well as your assessment of the examinee's communication skills.

An aspect of RTP writing that must be addressed is the use of proper grammar and well-constructed sentences. One issue that trips up many new lead graders is verb tense. Since the exam has already been completed prior to grading and evaluation, the past tense should always be used on the RTP when discussing what the examinee has written. Present and future tenses are appropriate when offering advice as to what the examinee should do either now or in the future. It is easy to slip, so a final proofreading should be carried out before sending the document on to the Associate Director.

Differences of opinion exist as to using an active or passive voice, and in using personal or indefinite references. As an example, the active/personal statement "You did not provide a description of sweet stout" would be written passively as "A description of sweet stout was omitted." In most cases, the chosen style is simply a matter of personal preference. Select a style that works for you, and be consistent.

Those who prefer the active voice often write using personal pronouns, while indefinite references are common with the passive voice. The active/personal approach can be more engaging and conversational, and thus sound more like coaching than lecturing. However, it can also seem accusatory when conveying negative information. Beware of this tendency if you select this approach.

The passive voice sounds more formal and impersonal, which may or may not be what is intended. It can provide a consistent matter-of-fact tone to the presentation, but it may also turn off people who are looking for more personal feedback. It also tends to generate longer prose, and the occasional awkward grammatical construct. If the passive/indefinite approach is used, beware of this tendency and seek ways to engage the reader.

Using complete sentences or even compound sentences also helps improve the overall quality of writing. Beware of excessive repetition of common words. Consulting a thesaurus will help in finding alternative words and phrases; online versions exist, such as www.thesaurus.com, or by using Word's *Look Up...* or *Synonyms* features (right-click on a word, then use the context-sensitive menu options).

When scoring style question responses, make sure that the examinee has addressed the essential elements of the style. This is much more important than matching nuances used in the Style Guideline descriptions. As an example when describing dry stout, the roasty character and creamy mouthfeel are important as well as the bitter chocolate character contributed by the use of roasted unmalted barley. These essential style attributes are vastly more relevant than correctly noting the ester level, since that is a minor element of this style.

The use of evocative terms when describing aroma and flavor elements adds value to answers. For example, vague phrases such as “moderate malt” do little to convey the essence of a beer. A description like “toasty, bready, cracker-like malt that finishes quite dry” is a much more expressive description.

When discussing the Tasting Section of scoresheets, it’s important to compare the perceptions of the examinee to those of the proctors. The completeness and quality of descriptions can be evaluated by noting whether or not the cues under each scoresheet section heading were addressed. For example, under the Aroma section, malt, hops, esters, and other aromatics should each be addressed. The “other aromatics” comments should be tailored to the style. For instance, it would not be unusual to detect DMS in the aroma of most Pils-based lager styles, so a comment about its presence or absence would be appropriate.

When shading the rank areas on page one of the RTP, it’s a good idea to review the Exam Grading Form (EGF) to see what rank levels the scores suggest. Question 1 determines the level of program knowledge. Relative levels of technical knowledge and style knowledge are calculated in the current EGF spreadsheet tool; make sure that the shaded ranks correspond to the actual scores granted. The level of communication skill pertains to the overall ability of the examinee to articulate relevant content, and also includes writing ability and even handwriting quality. For the Taste Portion, each shaded rank level should agree with the scores granted in the EGF. Remember to check addition on the scoresheets to make sure that all scores were totaled correctly. Finally, although scoring accuracy is calculated, it may be adjusted after considering the quality of proctor performance.

Best Practices for Writing RTPs

Many of the most common errors made while writing RTPs can be avoided by being consistent in assessing an exam, by using proper English, and by setting up and leveraging the many of the time-saving features of Microsoft Word.

- Mark comments and scores directly on the exams with a red pen or marker. This makes it easier to collect your comments when writing an RTP. The detailed feedback on individual questions can often be fleshed out from these comments.
- Avoid copying RTP #n to create RTP #n+1. Give personalized feedback, not generic lectures. Associate Directors and Exam Directors hate reading RTPs that all sound alike, which makes errors harder to catch. Corrective guidance for common errors can certainly be repeated from RTP to RTP, but it’s unlikely that two examinees will be so similar that the entire RTP can be reused. Some graders spend more time trying to edit one RTP to apply to another examinee than the time it would take to develop a new RTP. Whichever methods are used, the Lead Grader should make sure that all comments on the RTP apply to the subject examinee.
- Address what the examinee wrote, not what you would have written. There is more than one correct way to answer an essay question; allow the examinees to express themselves. If you find yourself giving the same feedback to all examinees, step back and examine whether you are looking for something reasonable or if you are trying to tell them how you would have answered the question. Don’t hold them to impossible standards, and don’t be a fault-finder.

- Recognize that full credit means that a question has been addressed competently without errors considering the time limit. Just because you can think of additional facts that have not been included doesn't mean that the answer shouldn't receive full credit. Focus on the important elements requested by the question.
- Consider the full response when assigning a point value. Don't deduct excessive points for errors. A good rule of thumb is to multiply the question score by 10 and compare it against the total exam score range for each judge rank. Then think whether or not that response is what you'd expect from a judge of that level (e.g., an 8 should show minimal National competence, while an 8.5 would be a solid National answer). Use this as a sanity check when deducting points.
- A skipped question is worth zero points, but be liberal in awarding partial credit if an examinee makes an honest attempt to answer all the questions. Partial credit neither should be given for non-responsive answers, nor should it be given for otherwise correct content that is unrelated to the question. However, if examinees make a legitimate attempt to provide information relevant to the question, then the first points awarded tend to come easier. For example—on the Classic Brewing Centers question—if someone just writes “Lambic” for Senne Valley and “Guinness” for Dublin, then they might earn two points. But if someone wrote a complete page-long response, would those two facts be worth that much? No.
- Remember that a person failing the exam doesn't need to be abused with a low score if they made an honest attempt at the exam. Reserve the low scores for those who obviously aren't trying. Use scoring to distinguish between those who tried but haven't mastered the material and those who clearly didn't try. The RTPs should always describe the exam as written. Clarify the reasons for failing exams, including describing lack of preparation and effort versus failing to adequately master the requisite subject matter.
- Try to determine if certain types of errors are being made consistently across multiple questions. This may indicate a gap in knowledge, or a misunderstanding of the exam format. This kind of feedback should be provided in addition to the analysis of individual questions. Often this type of feedback will help an examinee understand what type of additional studying is required. If such a trend is noticed, it is appropriate to refer to it either in the initial comments or in the wrap-up.
- Give a thorough assessment but don't be excessively wordy. Most exams can be adequately described in two written pages. If you are writing more than this, ask yourself whether you are adding value to your discussion. Are you being excessively repetitive? Could you rephrase your advice using fewer words? Are you quoting paragraphs when phrases would suffice? Are you talking more about yourself than the examinee?
- Don't use shorthand grader terminology exclusively to refer to a question (e.g., Troubleshooting Question), give some details on what the question was asking since the reader most likely is NOT an exam grader and isn't familiar with the question.
- Comments should match your scoring. Yes, we're trying to be encouraging but don't say that a person did a great job if they didn't. One approach is use standard terms for answer quality. A “poor” job on a question would be one that was less than 60. A “fair” answer would be one in the 60s, while a “good” answer is one that scores in the 70s. The “very good” scores are those in the 80s while “excellent” answers are those that score in the 90s. This is analogous to the terms for clarity that range from cloudy to brilliant.
- Pay attention to spelling, punctuation, grammar and style. Don't blindly accept any suggestions Word offers, but take a close look if it flags misspellings or common usage problems. Add brewing-specific terminology to your Word personal dictionary so that they won't be flagged. Enter properly

punctuated foreign terms, such as German brewing terms and cities. Using Word's *AutoCorrect – Replace Text As You Type* feature (Tools/AutoCorrect Options...) can make these substitutions automatically for you (e.g., Kölsch for Kolsch). Special symbols (degree symbols, em and en dashes, etc.) and characters (e.g., letters with umlauts) can be found in Word's *Insert\Symbol...* menu command.

- Punctuation belongs inside quotation marks. Use degree symbols when discussing temperatures. Be aware of the meaning of common homonyms (affect vs. effect, principal vs. principle, complement vs. compliment, etc.). When in doubt, check a Style and Usage Guide (such as Strunk & White's *Elements of Style*). Understand when to use a semicolon vs. a comma. Make sure your sentences agree in tense and number. Use conjunctions properly. Spell out numbers less than 10. Don't start sentences with numbers.
- Use Word's built-in bullet feature when assessing each question. To get the paragraphs to align properly, use *control-tab* instead of *tab* to indent the first sentence.
- Do not use the yellow highlighting tool when marking rank levels or suggested readings; use the Grey 25% tool. The yellow tool does not copy well. Use the text highlighting tool (part of the Formatting toolbar – *View/Toolbars*) rather than adjusting *Border and Shading* formatting.
- Do not use a full carriage return between paragraphs. Rather, use Word's *Format Paragraph* tool to set the "space after" to 6pt. This results in a more visually appealing format. Leave the sentences *Align Left*, not set to *Justify*. Be consistent in the number of spaces between sentences. With proportionally-spaced fonts, one space looks best but two spaces are also acceptable.

Working with the Second Grader

The second grader is your partner in reviewing the exams and assigning the scores. The skill level and experience of a second grader can vary greatly. Many will be new or very new to the program, but some can be more senior judges and graders (even Exam Directors and Associate Exam Directors) who wish to simply maintain their grading skills. Don't make unwarranted assumptions about your grading partner. If you don't know him or her, check their skill level, interest and goals for the process.

If you are working with a new grader or with an experienced grader who may be wishing to step up to a lead grader role, it is your responsibility to provide training in the process. It is helpful to have a quick discussion at the start of grading to establish communication and set expectations. While you should always provide feedback to your second grader at the end of the grading process, also provide immediate feedback when warranted. If you notice a problem, try to correct it quickly. Remember to provide your assessment of the second grader to the AD and ED at the end of the process, particularly if the person either is exceptionally good or needs additional development.

If you are working with a new grader, you may wish to suggest that you both grade one of the exams quickly and compare scores and major impressions. Think of it as a "calibration exam." This is a good leading indicator of potential scoring problems later. If you have major disagreements, see if you can get to the root cause of the problem so that the rest of the grading process can smoothly proceed.

Second graders often have varying levels of enthusiasm and time availability. Ask them if they want to provide input to the RTPs. If so, get bullet points for inclusion. If they have some experience as a grader, ask if they would like to prepare any of the RTPs. Ask if they understand the process or need any tips along the way. Ask if they want copies of the RTPs you produce, or if they want to wait until the exam set is closed. If they don't want to take on additional duties, don't be offended.

Good communications are the foundation of a solid grading experience. Touch base quickly and set expectations, particularly on intermediate milestones. Ask for commitments as to when grading will be

finished and then follow up if they miss the date. You are responsible for the grading team staying on schedule. The RTPs often take longer than expected to complete, so be sure to reserve adequate time to complete them. It's a good idea to rough them out immediately after (or even during) grading. Don't let open issues with any single exam keep you from working on the RTPs for the other examinees.

Email is the normal mode of communication between graders, but some may decide that a phone call will help in the reconciliation and consensus phase. If you do use a phone call, try to focus on only those exams where there are significant differences. Agree on the easy ones via email first so that the phone call can be used where an interactive discussion is required. If you disagree about a score for an exam, don't automatically assume that you're right. It's like judging beer; differences of opinion can exist. State your case, support it with facts, and argue using logic rather than emotion.

Working with the Associate Exam Director

The Associate Exam Director (AD) is the person who first reviews your RTPs and checks your scores before sending the package along to the Exam Director. Typically, it is the AD who gives the RTPs the most thorough review. These are senior graders who have seen a lot of exams.

The AD is likely to have multiple exam sets open at any given time. Your exam is not his or her only concern. Please treat their time as a valuable resource and use it wisely. Communicate with them when you have status information to share, but especially when the news is bad. Don't disappear! If you need help or can't complete the job, don't wait until the deadline has passed. Ask for help. If you are having difficulties with your second grader, escalate the problem to the AD and Exam Director (ED) so they can intervene.

Stay on schedule. Writing RTPs often takes longer than you originally estimate. Send along consensus scores when settled, even if the RTPs haven't been written yet. The AD might want to spot-check your grading using his or her copy of the exams. Send the RTPs to the AD as they are finished; you don't have to hold them all and send them in bulk. The AD has to read them one at a time anyway, so this parallel writing and reviewing process can speed up the closure of the set. This also allows the AD to give you feedback immediately if he or she notices some problem in the way you are preparing your RTPs.

When you send your RTPs and consensus scores to the AD and ED, clearly identify any exams where you are requesting a regrade. The AD and ED may or may not automatically regrade your exams, so highlight those needing a closer look. If you disagree on a score with your second grader, ask the AD for a second opinion. It is also helpful to identify the confidence level you have in your scoring. For example, if you and your second grader were in rapid and complete agreement, pass that comment along. If you took a long time to reach consensus, provide that information as well. This tells the AD how closely to look at your exam set.

Don't be discouraged if the AD suggests changes to your scoring and/or accompanying RTP. Many ADs conduct an independent scoring review of the exams to verify that the graders are scoring within acceptable ranges. With so many graders, each with their own subjective opinions, the grading by the AD and the ED helps provide consistency across all graders and all exams. Work with the AD and the ED to understand their reasons for making changes and learn from them just as new graders learn from lead graders.

If you want feedback on your RTPs, ask the AD and/or the ED. They should give it to you anyway, but it doesn't hurt to proactively request advice. They are likely to spend more time with you if they know their feedback is desired. You should receive the final RTPs from the ED with Word's *Track Changes* turned on once the exam is closed. This will show you what edits were made to your document. The AD is in a better position to give you more personalized feedback on the grading, scoring and RTP process.

Note that Word's *Insert Comment* feature may also be used by the AD or ED to insert messages as to why something was changed – these changes won't show in the final version that is printed for the examinee but will be there for you to read.

Conclusion

As a lead grader, you are a key player in the BJCP Exam Program. Lead grading is a rare opportunity to give something tangible back to the organization while providing the opportunity to groom a new generation of judges. It is an honor to be selected as a lead grader, and most BJCP judges will never get the chance to provide such an essential service. Remember, you are the gatekeeper for new judges who enter the Program.