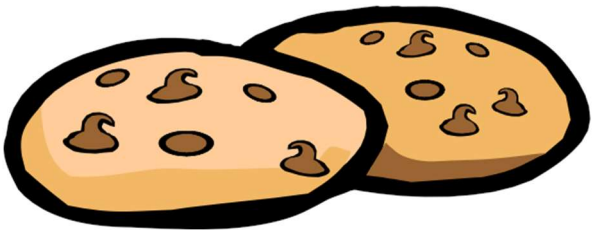
The Great American Chocolate Chip Cookie Taste Test
(AKA "Is everything in the rubric equally important to assess?")

This activity was adapted by Karin Hess from the original 1989 version created by Alysa Cummings and Karin Hess while working at the NJ Department of Education.

Context

The purpose of the activity is to introduce the concept of using a rubric with multiple and varied criteria as a tool for assessment. It provides teachers (and students) with experience in developing and using an assessment scoring tool and then transferring that learning to examine academic and 21st century-based rubrics (e.g., collaboration, self-direction) currently in use.



What This Activity Accomplishes

The universal appeal of chocolate chip cookies makes this both accessible and enjoyable for teachers and students alike. The development process helps participants to quantify the qualities of "good enough" which they later apply to classroom rubrics. This deeper examination of rubrics leads to two important key ideas when using rubrics with multiple criteria: (a) co-develop a common understanding of each criterion's intent before beginning to score student work; and (b) determine whether or not all criteria are equally important - some criteria could be weighted to better balance score points for performance reflecting deeper understanding (e.g., calculating v. problem-solving).

What Groups Will Do

Divide the large group into smaller groups of 4-6 people/group. Groups brainstorm the qualities of a good chocolate chip cookie, recording ideas on chart paper. Groups then decide which qualities to assess. If some ideas are related, they may want to combine several under one larger category (e.g., Physical features: texture + number of chips + size).

After identifying their categories (or criteria), groups define "performance" levels for each one. I suggest breaking down each criterion into 4 scoring levels, starting with descriptions of the "proficient" level. Once proficient cookie performance is defined by group consensus, the other three levels are described: from the lower end/novice level of the range to the optimum/expert level of performance. These descriptors are written into a group rubric (using chart paper or computer) and displayed for the whole group. Sometimes it's helpful to suggest adding specific examples of what they might be looking for at each level, such as the kind of chocolate used that reflects higher-to-lower quality. (Groups will not see the two cookies to be tested until the group rubric is completed.)

The leader hands out cookies A and B, but does not reveal brands. Groups use their rubrics to evaluate cookie A and cookie B in a taste test of two "blind" brands of chocolate chip cookies and scores are recorded.

After rating the two cookies, a spokesperson for each group must state conclusions using supporting data about the cookie rated the "best". Ratings are compared across groups about criteria used, how criteria are defined in those rubrics, and total scores given to each cookie to determine the top cookie. Do the final ratings accurately reflect the cookie with the highest quality? Is the number of chips more important than taste?

Time Required for Task

Approximately 45 – 60 minutes is needed for developing criteria, taste testing, and summarizing and comparing results.

Suggested Materials for Each Group

- Chart paper & markers
- Two chocolate chip cookie varieties to sample (I prefer using several in-store brands by the same company, otherwise "warm from the oven" will always win.)
- Napkins and two paper plates with "A" and "B" on them to keep things organized
- Rulers, Scales, or any other measuring devices (e.g., size or weight of cookies)

Possible Solutions

Results will vary depending on each group's criteria. It is important for the facilitator to compare criteria used by each group and descriptors for the same/similar criteria - just as two teachers might use different rubrics/criteria to score the same piece of writing and get different scores. Does anyone want to revise or weight any criteria?

Applying the Chocolate Chip Cookie Mindset

After the group sharing, ask: are all criteria equally important? What if a cookie has a high score in number of chips or size, but a low score in taste? Does that matter? What are the implications for rubrics we use in our classes or courses?

There are 5 general types of criteria used in most rubrics; process, form, and accuracy are almost always assessed. Creating knowledge and impact are the criteria that indicate FAR transfer - applying learning to novel contexts. When you create a rubric, keep these criterion types in mind. Consider whether some criteria are more essential to product quality and learning than others (e.g., taste v. size).

1. **Process criteria** – Did you follow the recipe (e.g., procedures for a science investigation; data collection or research methods; use of tools;/resources organization of a plan; correctly apply steps of a routine process)?
2. **Form criteria** – Is the size, texture, number of chips correct (e.g., did you apply correct formats and rules; complete product on time; use correct citation format; label diagrams correctly; organize parts; use required camera shots)?
3. **Accuracy of content criteria** – Is this a chocolate chip or another kind of cookie (e.g., is the answer/calculation correct; is the relationship accurately explained; is the concept demonstrated or accurately applied)?
4. **Knowledge production criteria** – What have you learned and can apply from making these cookies (e.g., going beyond finding a solution to gain new personal insights, raise new questions, and provide compelling evidence for claims or judgments made, such as in a self-reflection?
5. **Impact criteria** – How did it taste? (e.g., did the final product achieve its

intended purpose and provide evidence for claims, judgments, or a problem solved; did the solution work, change perspectives, or motivate action)

Analyze your group's cookie rubric criteria using the five types of criteria. Discuss whether some criteria are more important (e.g., taste), given the task. What weight or increased emphasis might you want to give to some criteria?

Next, analyze at your existing rubrics using the Cookie Mindset.

1. **Determine intent (what is uniquely being assessed), emphasis (most important), and potential evidence for each criterion.**
   - Which type of criteria are you using in this rubric?
   - Do the types of criteria match the critical aspects of the assessment task(s)? Do all criteria fully apply? Or only some?
   - Should some criteria get greater scoring emphasis? For example, at the start of the unit, you might place greater emphasis or only assess process skills (organization and procedures) when introducing what will be required in the culminating activity. By the end of the unit, impact criteria may have greater weight than data collection.

2. **Do a "DOK check" of your rubric criteria, assessment tasks, and unit lessons**:
   Does the wording of assessment prompts and rubric wording match the intended Depth of Knowledge* (DOK) expected to be observed in student work? Do the lessons build (scaffold) to the highest DOK expected by the end of the unit? Typically, each type of criterion will assess a different DOK level. **NEAR TRANSFER** = Applied as you were taught. **FAR TRANSFER** = Applied learning to a novel or unfamiliar situation.

   - **Process** – following the correct steps (a recipe) is usually only a DOK 1 level. If students are designing their own process, such as in deciding which variables to control and how to collect and organize data it would bump up to DOK 2 or DOK 3 depending on the task requirements. **NEAR TRANSFER**
   - **Form** – applying correct formats and rules (e.g., what makes a good oral or visual presentation) is usually only DOK 1 **NEAR TRANSFER**
   - **Accuracy of content** – locating/describing correct facts, details, definitions, or principles would generally be DOK 1; demonstrating use of an appropriate strategy for the situation, conceptual understanding, or accurate applications of concepts are generally DOK 2 **NEAR TRANSFER**
   - **Knowledge production** – providing support for new personal insights, judgments, and synthesized ideas can be DOK 3 or DOK 4 depending on the sources required and evidence used **FAR TRANSFER (A "stretch")**
   - **Impact** – evaluation of whether the final product achieved its intended purpose looks at coherence and effectiveness across criteria and generally, would be at DOK 3 or 4, since form, process, and accuracy are used to generate supporting/compelling evidence for ideas (creating new knowledge) for a particular purpose and audience **FAR TRANSFER (A "stretch")**

*For more on Depth of Knowledge, go to https://www.karin-hess.com/cognitive-rigor-and-dok.