



Evaluation Overview

Pilot Evaluation

Participants from a small set of pilot labs will complete a short survey both before and after engaging in the *Labs That Work... For Everyone* (LTW) professional development series. This survey evaluates participants' perceptions of the lab climate across four dimensions using the KINDER instrument, a brief adjectives-based climate measure, and measures of personal satisfaction, personal stress, and job satisfaction. We expect this survey to take participants no more than 15 minutes to complete.

During the program, after each episode, participants will be asked to provide rapid feedback through three questions: 1) if they enjoyed the scene/materials (yes/no), 2) if they found the scene/materials useful (yes/no), and 3) be given the opportunity to share any further thoughts with the evaluation team. These responses will be analyzed using standard structural equation, regression, and multi-level data modeling.

A selection of people from each participating lab—ideally at least one from each level of graduate student, postdoc, research scientist—comprising a sample of less than 75 participants will engage in a semi-structured interview so researchers can gain a better understanding of participants' experiences.

Interviews will explore which aspects of the modules and process participants found to be beneficial, and which aspects of the modules and process could be improved. They will cover information about participants' role in their labs, experience with LTW, observations and experiences in their labs during and after engaging in LTW, perceived benefits of LTW on both the individual and lab level, and an opportunity to provide open-ended feedback. Interviews will be conducted until we reach saturation of feedback themes. Interview data will be analyzed using reflexive thematic analysis to identify common and shared experiential categories.

Potential Efficacy Evaluation for LTW

After adjustments have been made to the program following the pilot, we anticipate rolling out the full LTW program across HHMI labs in successive cohorts. To evaluate the effectiveness of the program, we propose to conduct a fuller evaluation using the KINDER as a baseline advance assessment and as a post-program assessment. The Efficacy Evaluation will be conducted using a wait-list control design where labs will be divided into intervention and control groups. The control group will not receive the LTW development during the initial round of delivery and will participate in all evaluations and interviews. After the first round of LTW delivery is complete, the remaining, control portion of the labs will then receive the full LTW professional development course.

In such an efficacy evaluation, participants across HHMI labs would complete the KINDER survey on lab health and climate before and shortly after engaging in the *Labs that Work* program, and then again at a time between six and twelve months after completion of the program.

In addition to the scales evaluating the lab climate (KINDER), this evaluation would also measure personal satisfaction, personal stress, a brief adjectives-based climate measure, and job satisfaction as in the Pilot Evaluation (above), as well as a personal values articulation scale. Upon discussion with HHMI representatives, lab members' perceptions of the leadership behaviors of lab leaders may also be collected at all time points. During the program, after each episode, participants will be asked if they enjoyed the scene/materials, if they found the scene/materials useful, and be given the opportunity to share any further thoughts with the evaluation team. This data will be analyzed using standard structural equation, regression, and multi-level data modeling.

To complement the self-reported individual data with objective data about the labs' working practices and experiences, we will collect data on labs' attrition and turnover, labs' outcome/productivity (e.g., publications, researchers' placement), and labs' use of standard operating procedures (e.g., existence of data management plans, regular group meetings). These data will be collected via interviewing and gathering the needed information from lab coordinators or site administration.

Interviews will be conducted as outlined in the Pilot Evaluation (above). We will conduct these additional interviews until we are confident that no new themes beyond those collected in the Pilot Evaluation have emerged, and we expect to achieve that end with fewer interviews than in the Pilot Evaluation. With these additional interviewees, we may detect and pursue emerging themes for different groups of participants (e.g., lab leaders, graduate students).

In addition, the option will be provided to deliver the professional development in early- and late-stage waves with labs randomly selected into one of the groups providing the opportunity to conduct a controlled sample evaluation of the impact of the LTW development program.

