The PhET development process for creating and evaluating a simulation begins with the selection of the simulation design team consisting of between three to five individuals including a professional software developer, at least one content expert (scientist), a teacher and a student interface expert (education researcher). The design cycle starts with the content expert, teacher and student interface expert creating specific learning goals for the simulation. These learning goals draw from the existing research literature on conceptual difficulties with the specific content and based on our and other teacher’s experiences in the classroom.

A detailed initial layout for the simulation is created based on the learning goals and grade levels that are targeted. The simulation design follows the PhET Look and Feel\textsuperscript{1} which was originally created based on research on how students learn\textsuperscript{2} and from our first year of simulation interviews where we went through many iterations of interface features until we could successfully build a simulation that engaged students in scientist-like exploration. A complete discussion is provided in the pair of papers by Adams et al.\textsuperscript{3,4}. This PhET Look and Feel is a living document that slowly evolves based on new findings from our research on the simulations.

The next step is to show the initial layout to the developer to discuss feasibility, refine the interface and acquire a cost benefit analysis. Adjustments are made and the first version of the simulation is coded up. The team members communicate regularly to make any needed adjustments as the simulation takes shape. The simulation can be posted to our website and is labeled as “under construction” after extensive use by the team members and all members feel it’s clear, accurate and engaging. Student interviews are conducted at this stage with students who have the same background and academic preparation as the target audience for the simulation. These interviews always reveal interface weaknesses, resolve interface questions that were not agreed upon by the team, and often reveal pedagogically undesirable (and occasionally unexpected desirable) features and subtle programming bugs. Subsequent revisions are made, and if they are extensive further interviews are conducted with a new set of students. More recent interview results are finding much smaller problems than the interviews conducted on simulations that were written five years ago, indicating that our empirically developed design principles are working. After interviews establish that the desired engagement and learning is being achieved, the simulation is no longer marked as “under construction” on the website.

Each simulation is also used in the classroom by the teacher on the design team and often other teachers as well. Feedback from the teacher and any other observers are then used to identify possible issues that did not surface from the interviews such as features that allow unproductive playing. Once a simulation is used in a classroom setting where student use is observed and informally evaluated successfully the simulation is considered complete and receives the “checkmark” label on the web site. However, a simulation that has reached this stage is not set in stone forever.
Two to four times a year we engage in more formal rigorous studies of simulations in the classroom or through a series of interviews. These studies often reveal subtle changes that can improve student understanding of the concepts. In addition, we occasionally receive requests from users for new features or identification of subtle bugs that were missed during the above stages. Bugs are fixed and feature requests are logged and considered if the simulation is revisited.