

Assessing Affective Objectives for Communicative Visualizations

Elsie Lee-Robbins , and Eytan Adar 

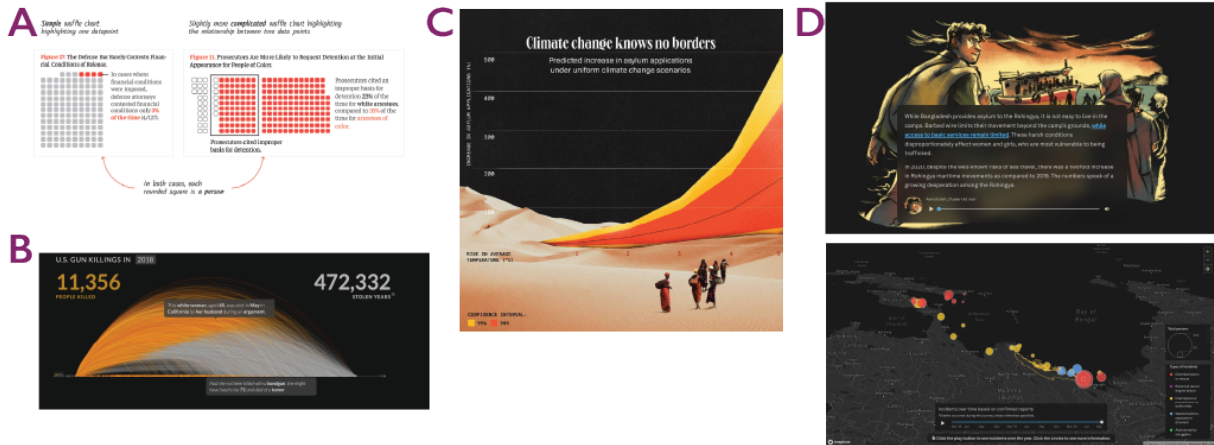


Fig. 1: Examples of several affective visualizations. (A) Freeing the US from its culture of detention: waffle chart, from Voila [84]. (B) US Gun Deaths, from Periscopic [63]. (C) Climate change knows no borders, from Gabrielle Merite [57]. (D) Two screenshots of Abandoned at Sea, from Kontinentalist in partnership with UN HCR [38]. The top screenshot shows an illustration of a refugee walking towards a boat, part of a narrative, and an audio file from Asmotulleh, a refugee. The bottom screenshot shows a map of refugees abandoned at sea.

Abstract—Using learning objectives to define designer intents for communicative visualizations can be a powerful design tool. Cognitive and affective objectives are concrete and specific, which can be translated to assessments when creating, evaluating, or comparing visualization ideas. However, while there are many well-validated assessments for cognitive objectives, affective objectives are uniquely challenging. It is easy to see if a visualization helps someone remember the number of patients in a clinic, but harder to observe the change in their attitudes around donations to a crisis. In this work, we define a set of criteria for selecting assessments—from education, advocacy, economics, health, and psychology—that align with affective objectives. We illustrate the use of the framework in a complex affective design task that combines personal narratives and visualizations. Our chosen assessments allow us to evaluate different designs in the context of our objectives and competing psychological theories.

Index Terms—Affective visualization, Visualization design, Assessment, Learning Objectives

1 INTRODUCTION

The use of learning objectives (LO) in the construction of communicative visualization [2, 43] can help designers better create and select visualizations [44]. Well-designed LOs have the benefit of being translatable to assessments [31]. Thus, a visualization designer can clearly state their intent—e.g., *the viewer will recall the unemployment levels in the US*—and create an appropriate assessment. For example, a simple multiple choice question: *What is the current unemployment level in the US? A) 5.1%, B) 4.3% or C) 4.6%*. By testing a viewer pre- and post-exposure to one or more visualizations, the designer can be more confident that their intents were achieved.

Learning objectives come in two primary flavors: cognitive and affective, and both can be used to define visualization intents (a third category, psychomotor, is harder to align). Cognitive objectives are easier to describe [3] using the standard **the viewer will [verb] [noun]** form. The verbs—recall, critique, summarize, etc.—are directly mappable to the cognitive changes the designer intends. Similarly, the nouns (specific facts, concepts, etc.) are more easily connected to data. The

clearer definitions of these constructs are a key part of why we can more easily construct *cognitive* assessments.

However, the intent of many communicative visualizations goes beyond the cognitive to the affective. For example, advocacy groups and organizations with a mission to improve the world often use data collected about humanitarian crises (see Figure 1). Their goals might be to raise awareness of the crisis and elicit responses from their audience: donate money to the cause, share the message with friends and family, or contact policymakers. Expressed as affective LOs: they aim to influence the audience’s appraisals, attitudes, or values [43].

In many ways, designing communicative visualizations for *cognitive* goals is more straightforward. There is significant guidance on their construction, significant research on their effectiveness, and there are rarely alternatives to visualizations. Cognitive goals are so entangled with the data itself that it is hard to argue that visualizations are not nearly always optimal for this task. With affective goals, there is far less guidance and research. Designing them well is important as the use of visualizations and data can actually backfire when design goals are affective. Impersonal charts or a focus on large population statistics, for example, can lead to numbing [74]. However, the answer to simply abandon visualizations is rarely the right one. Visualizations are often a response to a wicked design problem where multiple objectives—mixtures of cognitive and affective—must be balanced. In some situations designers can augment their designs with non-visualization features (either exclusively or in combination with

• Elsie Lee-Robbins was with the University of Michigan while completing this work, Eytan Adar is at the University of Michigan

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.202x.xxxxxxx



Fig. 2: (Left) A picture of Ruqio, a Somali woman, and her child in a hospital. Image from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) [81]. (Right) Data visualization of the population and location of Somalis facing food insecurity. Chart recreated, based on UN OCHA’s Humanitarian Response Plan Report [82].

visualizations). The key to making these decisions is to have good assessment tools. Intuition, even combined with specific design guidelines driven by well-studied theories, is unlikely to tell the designer which visualization design will work best or better.

The issue we tackle in this work is that the guidance on properly assessing *affective* LOs is very limited. Within education, affective assessments often focus on the class (e.g., did you like the class? do you want to take another math class? etc.) and suffer from various validity issues. Fortunately, experimental evaluation/measures from other domains (e.g., economics, advocacy, marketing, psychology, etc.) do offer alternatives. By restricting ourselves to assessments within the design process, we emphasize the need for timely (i.e., scalable and easy to run), actionable (that allow for contrasting design alternatives), and well-validated assessments. Our goal in this work is to articulate these constraints as guidelines for selecting assessments. Through the guidelines, we can dismiss assessment approaches where constructs or other implementation challenges make them unsuitable for our context. For example, certain political constructs are too vague, psychological scales may be unreliable, and advocacy evaluations can require time frames that greatly exceed a visualization design cycle. On the other hand, we can reclaim instruments which may be unreliable in certain environments (e.g., self-reports in a classroom settings) because they are reasonable in our setting.

We demonstrate how to utilize different affective assessments within a realistic case study. In the example, the designer has a mix of affective objectives (e.g., *believe in the importance of the humanitarian efforts*, *act to donate*, etc.). To achieve these, the designer can apply a mix of rhetorical and design strategies. They may leverage a logical argument (e.g., the rhetorical *logos*) and use statistics, numbers, and charts to draw a compelling picture of an issue by presenting data as evidence. Numbers and statistics can be presented as facts, wielded as reasons to support a persuasive argument. Data visualization is a tool that can effectively show trends, patterns, and the scope of problems (Figure 2, right). Alternatively, they may emphasize an emotional appeal (the rhetorical *pathos*) by showing photographs, audio, videos, and stories about individuals. For example, photography is often the most popular choice for certain types of advocacy communication [67] but other types of illustration and visual art are also common (Figure 2, left). Using the LO formulation and assessment approaches, we demonstrate how different communicative alternatives can be compared.

This work contributes guidance on how affective objectives can be measured. Specifically, we describe the unique constraints of the assessment of affective visualization and a set of evaluation suggestions for a wide range of affective objectives. We offer a specific example of how proper assessment can inform a designer challenged by multiple objectives and conflicting research evidence.

2 BACKGROUND

2.1 Affective Visualizations

Affective visualizations have gained more attention as researchers consider non-cognitive intents [41, 43]. The term *affective* has several definitions in this community. In this work, we utilize the definition of Lee-Robbins and Adar [43] that treats affective goals as those that

aim to influence appraisals, attitudes, values, or value systems (broadly, *beliefs*). There are many types of beliefs that can be considered under this definition, and a clear articulation is challenging. However, for our purposes, we consider those constructs where (a) they can be learned or changed, and (b) where different people might hold different stances (e.g., positive or negative) on those beliefs. Defining these by exclusion, we *do not consider* affective targets that are primitive (non-reliant on other beliefs) and for which there is 100% consensus (everyone agrees) [69] (e.g., beliefs that are “psychologically incontrovertible”). Succinctly, these are ideas where not everyone agrees on the same thing and there is an opportunity to change a person’s mind.

Affective visualizations (see Figure 1) are visualizations created to achieve one of these goals. A specific category for affective visualizations is anthropographics, which have been defined as “*visualizations that represent data about people in a way intended to promote prosocial feelings (e.g. compassion or empathy) or prosocial behavior (e.g., donating or helping)*” [58]. Affective communicative visualizations can exist for many reasons, but they are often applied as a form of advocacy—pushing the viewer towards having some negative (or positive) view of some belief or assertion and behaving accordingly. Because advocacy often involves various rhetorical strategies, this area offers many examples of evaluation approaches of affective visualizations.

Advocacy-focused goals are a useful representation of affective communicative goals (and consequently affective visualization goals). At their core, advocacy aims to help other people. For example, an advocacy organization can launch advertising campaigns to raise awareness about an issue. Often, they organize fundraising drives to get donations for their cause. Non-visualization approaches may be to present testimonials, stories, and in-depth information about individuals. Individual photographs (e.g., of a person impacted) are often effective tools. People’s stories and photographs capture attention and can evoke an emotional connection (a *rhetorical pathos* strategy) [9].

An alternative approach, which we focus on, is to make a more data-driven appeal to the audience [29, 79]. Nonprofit and government organizations have increasingly used data visualizations in advocacy reports in recent decades [67]. There are several reasons why advocacy groups might want to present data visualizations. First, data could give advocacy groups credibility and trustworthiness [67] (an *ethos* rhetorical strategy). Data provides a logical argument and ‘evidence’ to support claims. Second, some designers perceive data visualizations to be persuasive [62, 67] (a *logos* strategy). Third, data visualizations could simplify and distill complex information and make it accessible to a wider audience [29].

Outside of advocacy, we see examples of affective visualizations in marketing, political propaganda, or health communication. In all these scenarios visualizations act as rhetorical objects, often targeting both cognitive and affective channels. Their objective is to convince the viewer of something and have them act accordingly—to buy, to vote, to take care of their health. Arguably, most visualizations, even those that aspire for neutrality or unbiased presentation (e.g., a visualization in a newspaper article or academic presentation) have some implicit affective objectives, often attempting to convince the viewer of the ‘correctness’ of an argument [43]. Assessing whether these affective objectives have been met is a key focus of this work.

2.2 Affective Learning Objectives

In this work, we leverage our previously-defined LO taxonomy [43]. The taxonomy offers a classification of verbs and nouns in the standard framing: *the viewer will [verb] [noun]*. This differs in a number of ways from the traditional affective objective taxonomy of Krathwol et al. [39], which lacked concrete structure for the ‘nouns’ and did consider a broader set of objectives beyond those in a classroom setting. It is also worth noting that we take a broad understanding of ‘affective’ to include not just emotions and feelings. In the case of cognitive objectives, *perception* (e.g., reading the height of a bar in a bar chart) is necessary but not sufficient for many objectives (e.g., recalling something’s value). Similarly, for affective objectives, an emotional response is often insufficient (and may, or may not, be necessary). Rather, a designer will seek to encourage changes in beliefs, attitudes, or values

that persist beyond the initial emotional reaction. More critically, an emotion-driven strategy (i.e., pathos) may not also be appropriate to achieve an affective rhetorical objective.

Figure 3 (left) summarizes the noun/verb grid for the taxonomy with examples. We note a few specific attributes of the taxonomy in that they impact our assessment choices. The verbs (rows) focus on what we want the viewer to be able to do after being exposed to the visualization.

This affective objective structure defines the space for which we would like to identify assessments. As with most ‘learning’ situations, a designer might have multiple objectives that are a mix of affective and cognitive and may be in opposition to each other as part of a wicked design problem [13]. This will necessarily impact design choices and implicate assessment strategies. Additionally, as with cognitive objective frameworks [3], verbs and nouns are only weakly hierarchical [32]. For example, some verbs (e.g., *Position*) may depend on others (e.g., *Observe*). However, it is not necessary to achieve a *Connect* objective before targeting a *Behave* one. Again, this will have an impact on how assessments are structured, as it is possible that achieving a ‘high-level’ objective such as *Behave* or *Connect* does not mean the *Strengthen* objective has been met.

2.3 Affective Assessment in Education

In theory, the advantage of creating a well-defined set of learning objectives should mean that these are translatable to assessments that can cover a wide range of topics. In the case of cognitive learning objectives, this has certainly happened as instructors have created questions, tests, and rubrics for all types and topics. Unfortunately, the same cannot be said for affective objectives.

Taxonomies for affective objectives have evolved over time, but nowhere near the rate of other objective types [39, 54]. Unlike Bloom’s original cognitive taxonomy, which has seen continuous refinement and updating [3], affective objectives have remained more fixed. In part, this is due to the difficulty of assessment. The 1956 form of the taxonomy which includes a hierarchy of verbs (*receive, respond, value, organize, and characterizing*) but no principled model of the nouns to which they should be applied. The bulk of examples from affective education target a specific set of objectives that are often about the act of learning itself (i.e., ‘the student will appreciate math’, ‘the student will register for more English classes’) or classroom behaviors (‘the student will not bully others’) [4, 18].

A complete discussion of why education has shied away from broader engagement with affective objectives is beyond the scope of this paper. However, we consider a few key aspects of this legacy. First, there are often ethical concerns for affective objectives. That these may reflect advocacy or even brainwashing [54] and a deviation from the ‘knowledge credentialing’ focus was inappropriate. Outside the classroom, as communicative visualizations often are, this concern is more limited¹.

A second issue with affective objectives in educational contexts is that they are often assessed anonymously and/or without grading [53] and thus are unreliable. Although one could reasonably construct a test to assess math knowledge or a rubric to evaluate a written document, affective objectives are much more challenging. In the case of self-reports, which is the predominant educational assessment mechanism, the test taker often recognizes the social- and teacher-desirable answer and will respond accordingly. In the context of communicative visualizations outside the class, these issues may be less important. The viewers are not incentivized to produce socially-desirable responses and more incentive-compatible and elicitation protocols can be used.

Finally, classroom assessments are often primarily aimed at evaluating individual students rather than determining programmatic value (i.e., how much the teacher, textbook, activity, or module helped). That is, the goal is to determine an individual student’s performance either relative to others in the class or to some baseline. In our context, assessments are entirely focused on the program and can easily produce and evaluate alternatives (e.g., ‘does the visualization help achieve the

objective?’ and ‘which visualization is better?’). It is much easier to produce alternative designs and find an audience for evaluation. This will make many other types of assessments available to us.

3 SELECTING AND CONSTRUCTING ASSESSMENTS

In identifying assessments for affective visualizations, it is critical to ask: what is unique in the *design process* of communicative visualizations? Primarily, we are interested in those time-bound activities in which a designer produces an artifact that contains visualizations. These visualizations may be static elements for an article, an infographic, a video, a presentation, or interactive treatments for the web. In most projects, the designer will have some mix of cognitive and affective objectives—some of which may be in conflict with others or be prioritized differently. Visualization design under these constraints is often a *wicked design* problem [13]. A solution may never be perfect or optimal, but may be produced as part of some tradeoffs. ‘Mastery’ of any particular objective may never be achieved. In most cases, the designer is seeking to do the best they can towards satisfying their objectives under constraints. Assessments offer a mechanism by which they can determine progress.

3.1 What’s better?

In its simplest form, an assessment simply resolves to some operationalized measure of learning, or in our case, belief. In many situations, this treats the visualization as an intervention where the same question can be asked before and after exposure (during or after showing the visualization): How important is this topic before I showed you the data? and how important do you believe it is after? Naturally, the designer might have many different objectives, each with their own indicator. Assuming that no visualization ‘dominates’ all objectives, this would mean making a tradeoff of which objectives are most important.

Not all assessments require both pre- and post-tests. When comparing alternatives, the designer may be satisfied with showing different people (e.g., half the participant population) the different options and comparing their responses². The advantage of visualization design is that design variations are often a natural output of the design process. During the design process, the designer can produce alternative designs that they intuitively believe will better serve their learning objectives: transforming the data, changing encodings, adding visual embellishments, creating specific annotations, modifying colors, etc. Thus, the act of design will invariably produce alternative forms.

We note that a between-subjects strategy is likely preferable for the assessments we describe here. That is, each participant only sees one version of the design. Clearly, exposing a viewer to multiple visualizations in a row might create affective priming from the first visualization they see, impacting the results for the second.

As we describe below, each domain (education, marketing, politics, non-profit advocacy) will offer assessment options. However, not all are suitable for design work. In comparing options, we focus on approaches that are timely, actionable, and valid in the context of design. In no case are these ‘absolute’ rules as there will be scenarios that require alternative approaches.

Timely—Assessments should be able to influence design decisions and be integrated into the design practice (ideally, they should be cheap and scalable). First, most visualization design activities are very time-constrained (i.e., deadline driven). Second, once deployed, the impacts of most affective visualization are likely short term (i.e., we would not expect a change in someone’s value system 3 months after seeing a single visualization). Although some projects may run for extended periods and have many iterations of develop-deploy, most do not. Our preference will be for measurements that can be executed rapidly without significant involvement from the designer. For example, assessments that track longitudinal behavior and ‘grading’ through a rubric or bio-physical measures [6] are less desirable. Similarly,

¹Though this does not mean that affective objectives are uncontroversial in the information visualization community (see [43] for a broader discussion).

²A more statistically sophisticated version of this would be the difference-in-difference (DID) [1] which could control for differences in the population looking at visualization 1 and those in visualization 2

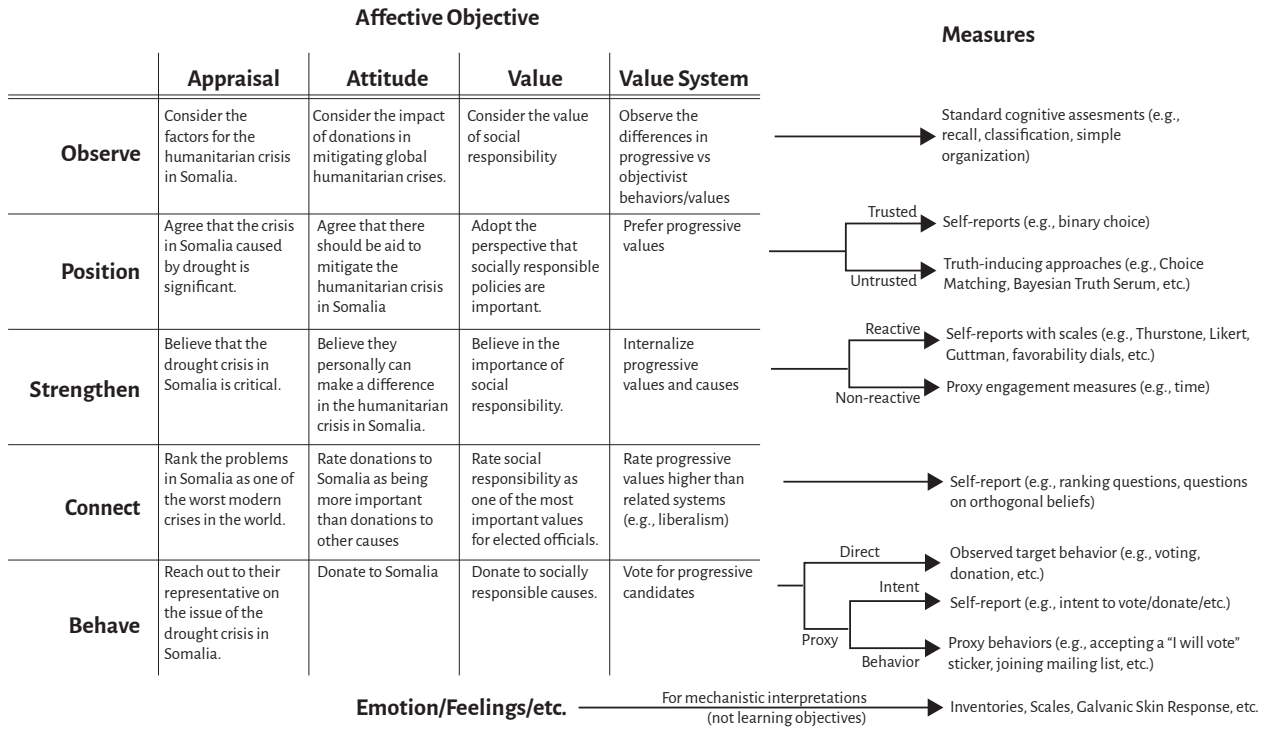


Fig. 3: Structure for affective learning objectives in communicative visualizations with possible measures. Each cell in the table (left) provide an example objective.

assessments that are unfamiliar are likely inappropriate (e.g., repertory grids [45] where participants create their own scales).

Actionable—As we have noted above, our goal is that assessments are useful for program-focused measures. We are much less interested in the impact of a visualization on individual viewers. That said, if we care about deploying different visualizations to different communities, we may consider groups of viewers. Ideally, the measure used should be sensitive enough so that we can isolate differences without having to recruit many participants. The differences (e.g., mean change in donations, difference in total donation, etc.) should allow the designer to make a decision of which visualization is better for their objectives. Measures that require huge populations will be less desirable. For example, randomized response methods use a coin flip to tell participants to either provide a fixed answer (e.g., ‘yes’) or answer truthfully [34, 86]³.

Valid—Clearly, we would like assessments to satisfy all the constraints of good experiments: they should target the right population and should be reliable. They should have construct-validity (match the objective’s construct), content-validity (sufficiently capture the objective), and predictive-validity (can we make an inference on the success of the objective from the measure?). None of these are unique to our domain. However, in the design case we leverage certain common features (e.g., access to multiple designs and potential for population randomization as a control).

3.2 Assessments by Verb

Having defined a few simple criteria and identifying the properties of our environment, we can now consider various evaluation approaches. We will work to map these to the specific objective structure (see Figure 3), roughly from the top to bottom row.

An important thing to note is that while assessment methods are often consistent along row (i.e., LO verb), the *language* or *scope* of the assessment may vary by the knowledge dimension (i.e., columns of

³This provides plausible deniability to individual participants (around sensitive topics) and increases truthful answers. However, many participants are needed to remove the biased answers.

Figure 3). These categories are Appraisal, Attitude, Value and Value System. An *Appraisal* is a value judgment of a data fact (e.g., ‘the drought in Somalia is devastating a huge number of Somalis’). An *Attitude* is a subjective opinion of an appraisal. For example, ‘the drought in Somalia is among the worst crises today.’ *Values* are more persistent beliefs that impact how the viewer interprets many situations. For example, *equity* is a specific value through which we can view many situations such as humanitarian aid, hiring decisions, etc. Finally, *Value systems* are groupings of values (e.g. a *progressive* or *conservative* value system). In describing the assessment strategies, we also discuss how they can be adapted to these categories.

3.2.1 Assessments for *Observe*

Observe, the most basic objective, aims for the viewer to ‘see’ a space of possible beliefs (or appraisals, attitudes, etc.) exist. The target of this verb may be ‘free-standing’ (e.g., ‘greenhouse gasses are bad for the environment’). As often, the target will be a spectrum of two things held in opposition (i.e., a debate). For example, one might hold the belief that ‘eliminating greenhouse gasses is the way to eliminate climate change’ versus ‘eliminating aerosols is the way to eliminate climate change.’ This space of views need not be a one-dimensional spectrum (e.g., there are many causes of climate change and which is most important can be debated).

Observe objectives, whether targeted at appraisals, attitudes, values or value systems—are likely to be the simplest to measure using more standard assessment approaches (e.g., [32]). In this category, the designer wants the viewer to roughly model the belief space in question. For example, there is an argument over the factors that are leading to the crisis in Somalia (appraisal); or that individuals might believe in social responsibility through individual or government interventions (value). Assessments that can check recall, classification, or simple organization would likely work well here. For example, we could ask *select all factors from the list that are impacting the humanitarian crisis in Somalia with a list of true and distractor factors*. If the objective requires a deeper analysis, the question might be about properties or contrasts. For example, *for each action in a list, indicate if they are the*

actions of someone who believes in individual social-responsibility or governmental social-responsibility.

3.2.2 Assessments for *Position*

Position objectives go a step further beyond seeing the space of possible opinions one could take. These objectives specifically seek to place the viewer into a point in this space or that one point in this space is better than another (e.g., convincing the viewer to sit closer to the ‘greenhouse gas’ side of the argument than the ‘aerosols’).

Within the position level, designs can be assessed through measures of attitude and attitude change. Roughly, we would like the viewer to prefer one position over another. Many measures in this space are based on self-reports. Agreement or disagreement on whether one believes in a statement of appraisal, attitude, etc. may provide a signal of the viewer’s placement.

Attitudes can often be measured directly or indirectly through self-reports [78]. For example, a viewer can be asked directly if they agree that some position is favorable (e.g., *Recycling is good*). They can also be asked indirectly by making judgments about the behaviors of others (e.g., *In this scenario, Sally threw her can in the trash, do you agree with this behavior?*).

Recall that a specific concern for visualization design is sufficient sensitivity, so we can see a result with fewer viewers. The formulation of survey statements can have a significant impact on the results. For example, agreement with the statements that *we should allow speeches against democracy* versus *should not forbid speeches against democracy* (ostensibly measuring the same attitude) results in different levels of agreement: 21% vs 39% [71]. When comparing two visualizations, both statement forms may yield a reasonable ordering, but there is some risk if this is the only thing tested. Adding additional questions or survey instruments to test the same instrument may boost our confidence and mitigate the effects of a bad survey choice, while not adding too much burden on the participating viewer.

When self-reports are still untrustworthy, potentially for opinions that may have high social-undesirability, an alternative is to leverage the truth inducing approaches from psychology (e.g., error choice method [33]) or economics [16]. Methods such as Choice Matching [20], Bayesian Truth Serum [66], and Bayesian Markets [5]. Such methods often ask about the participants view’s but also to guess or bet on what other people said. A correct guess may yield an additional reward. Such approaches are more incentive-compatible, but may require a certain sophistication and rationality in the respondents.

3.2.3 Assessments for *Strengthen*

Strengthen objectives build on *position* even further by having the viewer more strongly ‘believe’ or ‘disbelieve’ in a particular viewpoint. Assessments that target the strength of a specific belief are often based on self-reports. There are various psychometric measures that have been extensively validated and can produce an indication of how strongly someone believes something. These include instruments such as Thurstone, Likert, Guttman, or Semantic Differential (see [17, 24, 40, 55]).

Figure 4 illustrates examples of scales. Some, like the Thurstone scale, require collecting favorability ratings by experts if constructed from scratch. However, there are many validated versions for specific topics that can be used directly or adapted (e.g., around social activism [80]). Some scales are more familiar (e.g., Likert) but may not be as easy to combine into one number. Scales such as the Guttman Scalogram are designed hierarchically so that agreement with one statement implies agreement with others. Semantic differential scales have the advantage of evaluating an opinion based on multiple criteria. In addition to written visual scales, visual scales such as the Self Assessment Manikin [12] allow a participant to select a cartoon image representing their ‘feelings’ on a subject.

In addition to the various scales above, there are a number of non-reactive and unobtrusive measures (see [87]) that can be used to gauge opinion (both position and strength). For example, in an interactive system one could log engagement or time used as a proxy measure for an affective objective (e.g., see [65] for the case of a MOOC). Other

Likert

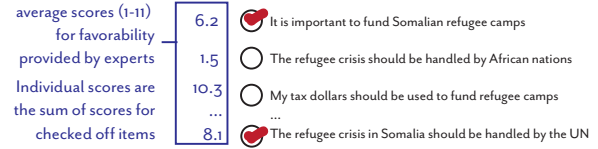
instructions: rate your agreement for each statement

The humanitarian crisis in Somalia needs to be immediately addressed



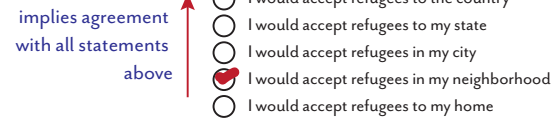
Thurstone

instructions: mark all statements you agree with



Guttman Scalogram

instructions: mark the statement that best represents your views



Semantic Differential Scale

instructions: best express your impression for each of the following



Self Assessment Manikin

instructions: circle the best representation of your happiness

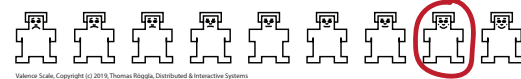


Fig. 4: Examples of different scale types

proxy measures might be to provide the viewer with an opportunity to share, save, ‘like’ or comment on a visualization. While most measures we have considered thus far are used in pre- and/or post-evaluation, instruments such as debate dials can be used when there is temporality to the presentation (e.g., the a video presentation of visualizations). Viewers can slide the dial in one direction or the other indicating (un)favorability while observing the material [72].

3.2.4 Assessments for *Connect*

Connect objectives are those relating to situations where the viewer needs to contrast two or more beliefs that may not be in opposition to each other. For example, the designer might want to acknowledge that climate change and economic inequality are two pressing concerns for humanity, but that a point in the climate change space (e.g., ‘greenhouse gas-caused climate change’ is more important than ‘globalization’s impact on inequality’). These four objective types—observe, position, strengthen, and connect—are all ‘internal’ objectives for the viewer.

Connect objectives can be studied similarly to Strengthen and Position objectives. They simply involve additional instruments that consider additional orthogonal beliefs. For example, we might consider all the humanitarian crises and test the viewer’s attitude towards them in relation to some target (e.g., a crisis in Somalia). As with Position objectives, Connect objectives might use ranking measures to determine how important certain issues are to an individual.

Advocacy evaluation in particular often considers priorities as a key measure [68, 88]. In evaluating these efforts, participants are asked

questions of the form: *how seriously (do you / does your community) take each of the following issues? or how likely would it be for you to support state funding towards each?* Such metrics may be adaptable to other affective objective scenarios. One specific constraint of advocacy, which may implicate visualizations used for advocacy, is that efforts of strategic communication may be targeted at very specific people (e.g., politicians who vote a certain way). In such cases, assessment may involve a narrower population of ‘viewers’ (e.g., interviews with bellwethers or legislative staff [68]).

3.2.5 Assessments for Behavior

A final objective category, *Behave*, is the externalization of beliefs into behaviors (e.g., donating, voting, publicly advocating for, etc.). Ultimately, many affective objectives are about behavior. A behavior goal is that people should *do* something based on a belief–vote, call their representative, donate, recycle, stop smoking, etc. In most cases, it is difficult to observe a behavior during an assessment of the design. That is, it is not always possible to provide an opportunity for the viewer of the visualization to demonstrate the behavior. In other cases, the desired outcome is that people stop doing some behavior (e.g., stopping to smoke). Observing or measuring this kind of behavior may present its own set of challenges. Due to the challenges of direct observation, many rely on behavioral proxies such as self-reports to measure *intention to behave*.

Behavioral intentions *can* predict behavior [28] assuming certain conditions are met. These include: specificity (*I intend to donate \$5 to X*); stability of the intention (how likely is it my intention to hold in the future?); and volitional control (how much control over the behavior does the person actually have). Clearly, not all behaviors align well in this way. The result is that we often see a weak correlation between self-reports and behavior [22]. Again, because we are often comparing two or more visualization designs, knowing which leads to a higher intention to behave is better than no measure. However, it may not indicate the same performance once the visualization is deployed.

As with other objectives, unobtrusive measures might be used as proxy measures for behavior [87]. For example, accepting a recycling sticker may indicate an increased chance of recycling. As with other measures, unobtrusive measures require validation and therefore may not be suitable for all design tasks or objectives.

Another option for measuring behavior is to create opportunities within the assessment framework that allow for the demonstration of behavior. As we illustrate in the following case study, donation is one such measure. In addition to being closely aligned with the actual objective of interest (i.e., donation), this approach is incentive-compatible and preferable to self-reports. Not all behaviors can be simulated this way, although proxy behaviors (e.g., donation to a candidate) are correlated to other behaviors or beliefs (e.g., voting behavior or policy preferences [10]).

3.3 Measuring Emotions and Other Theories

An astute reader may notice that despite our focus on ‘affective’—which is often associated with *emotion*—neither our taxonomy nor assessment approaches focus on emotion as a central concern. In the context of *research*, where mechanistic explanations are important, there are many instruments to assess emotional response [25, 56]. Emotional responses fall into a broader class of constructs that relate to affective learning, often within a theoretical framing in psychology, persuasion theory, marketing, and so on (e.g., [8, 24, 27, 46, 61, 89, 91]). For example, the Elaboration Likelihood Model (ELM) [64] looks at attitude change through the ‘route’ by which that change can be induced (i.e., a central, deep reasoning route, or peripheral, more emotional or intuitive assessment). The model also provides an argument for what constitutes a strong (eliciting mostly positive reactions when thought about deeply) or weak message (mostly negative) in terms of elaboration.

Designers could use knowledge of emotion’s role in affective learning and apply that to their own design. For example, a designer may know that rage can be manipulated to induce donation [15] and may shy away from the colors that induce positive reactions and select those that induce disgust [36]. However, the impact of a color choice in a

specific design on emotion and then donation may be impossible for a designer to practically determine.

In some ways, measuring emotion would simplify the designer’s task. However, as we noted earlier, our model of affective objectives is that they should target things that can be learned, and thus retained. Emotions, although they influence behavior and learning, are often fleeting (e.g., [30]). Beliefs, when they can be influenced, tend to be persistent (e.g., [73]). A particular focus on emotion may not provide the timely, actionable, and valid assessments that would benefit design.

4 CASE STUDY

In this case study, we showcase how to use the guidance in this paper to select and implement assessments based on affective learning objectives. We use the topic of a humanitarian crisis in Somalia, as it is a topic that most people do not have extensive prior knowledge about and has impact for affective learning objectives around attitudes, beliefs, and behaviors. Somalia experienced internal displacement and food insecurity as a result of a drought that affected 7 million people. We illustrate how learning objectives and assessments can be used in a real-world context by simulating our roles as designers working in collaboration with UN OCHA (the United Nations Office for the Coordination of Humanitarian Affairs).

For this project, we are tasked with building an advocacy video ⁴. In this hypothetical design, UN OCHA has multiple documents about the crisis that are the ‘seed’ to the videos. Some focus on individual stories [81] and other on data [82]. The target audience of this design project is members of the public who are unaware of the crisis.

Our design process follows a fairly conventional strategy: (1) identifying the goals for the project, articulated as learning objectives; (2) create several designs that could reasonably achieve our goals; and (3) choose an assessment for each goal by using the framework and guidelines in Section 3. For this case study we demonstrate the implementation of the assessments through a crowdsourced user study on Prolific. Finally, we discuss the results of the survey and how we could use the information to inform our design process.

4.1 Affective learning objectives

In our design process, we created these objectives. We infer these based on a close reading of materials from UN OCHA (e.g., [81, 82]).

- LO-A: The viewer will observe the humanitarian crisis in Somalia.
- LO-B: The viewer will believe that the crisis in Somalia is important.
- LO-C: The viewer will believe that they are able to personally make a difference in helping people in Somalia.
- LO-D: The viewer will believe that they have a moral responsibility to help people in Somalia.
- LO-E: The viewer will practice social responsibility and donate money to the Somalia Humanitarian Response Plan.

These objectives capture a range of LOs from our taxonomy (observe, strengthen, and behave).

4.2 Visualization designs

As designers, we created our artifact in the form of a short video that builds on existing UN OCHA materials. The existing materials give us many design assets (text, visualizations, maps, photographs, etc.). The fund’s annual report (Humanitarian Response Plan [82]) significantly emphasizes data and visualizations. The report is published each year by UN OCHA and provides detailed information on the crisis and data on relief efforts (111 pages in the 2023 version we used). Critically, it contains many data visualizations on the current situation in Somalia, including maps, bar charts, donut charts, bubble/area charts, icon arrays, stacked bar charts, and line charts.

An alternative document is a long-form humanitarian essay with stories about individual Somalis and Somali families [81]. The essay,

⁴While this is a simulated case study, we used reports and assets published by UN OCHA to create our materials. Some materials and data originally from a research project about data visualizations and statistical numbing.

titled “Somalia: Hope fades as famine looms”, had numerous high-quality photographs, a significant amount of written content to use, and the content was compelling and emotional. The essay had a link to a donation site, suggesting that this would be effective content to achieve our affective goals.

We identified three main topics of overlapping content in the Humanitarian Response Plan and the essay (acute food insecurity, internal displacement, and response efforts). These formed the core of the produced videos. In addition, we used written content from the foreword of the report, which summarized key points about the data. This became video voice-over when data visualizations were shown.

The challenge for us in achieving the LOs is that we have alternative starting points *and* conflicting information on what combination of visualizations, data, and narratives will work best.

Data-Focused Content—There is conventional wisdom in design [67] and research [52] that visualizations can shape belief. This can be achieved through a focus on data [62] or in meta-visualizations (e.g., ‘fundraising thermometers’ [35]). Given this, a natural video format would leverage the 2023 report’s visualizations and text. In cases where we needed to supplement figures from the report (e.g., the text made a strong point but there was no visualization), we leveraged visualizations from news outlets (e.g., The Economist [21]). The visualizations were across sources were well constructed and clearly communicated the need for humanitarian aid in Somalia. They made a logical argument about why a person should care about, and donate towards, the humanitarian crisis.

Human-Focused Content—As designers we are aware that data visualizations can also be less effective in increasing prosocial behaviors than just highlighting the people in need. Several studies point out possible problems in focusing on data. For example, the identifiable victim effect is the effect where empathy is highest for an identifiable, determined, or known person [37, 76, 85]. Empathy is highest for a single individual, and every additional individual is valued at a lesser value than the one before. Even just the move from one person in need to two people in need results in compassion fade [85]. This trend of compassion fade continues in even larger numbers with psychic numbing, the phenomenon of people’s decreasing empathy reactions to large numbers of suffering people [49, 75].

Another drawback to showing the entire scope of the problem is that it can induce pseudo-inefficiency, where people think that their potential action will not be useful since it is just a drop in the bucket [83].

Photojournalism portrays humans in their work [47, 90]. Faces remind us that the data set is made up of real humans; designers should “show what the data are about” [70]. This strategy is powerful because it is easier to form a mental image of and relate to a single person compared to an abstract dataset of people. An example of this is the photograph of Alan Kurdi, which sparked empathy for Syrian refugees that had not been elicited in previous news coverage [23].

Considering the disadvantages of compassion fade and pseudo-inefficiency and advantages of photographs, the 2022 article would potentially be a better model for the video. Thus, our second video design focused on stories and photographs of impacted Somalis. We use both the visual content (photographs) and written content in this video. The written content became the narrative voice-over in the video while the photographs are being shown.

Balanced Content—Because there may be benefits to both data-focused and human-focused approaches, a reasonable design hypothesis is that the combination is the ‘best of both worlds.’ To test this, we built a third video that attempts to integrate content from both videos.

One design strategy is to redesign the data visualizations to incorporate elements of *anthropographics*. Anthropographics are data visualizations that are designed to evoke feelings in the audience, such as empathy [58]. There are various design strategies of how and what to visualize when visualizing data about people [11, 58]. Past work has identified seven dimensions of anthropomorphic design: granularity, specificity, coverage, authenticity, realism, physicality, and situatedness [58]. Various data visualization designs in this space have been investigated to see their effect on prosocial feelings and behaviors [11, 14, 48, 59, 60]. However, these designs have been suggested to

only have a negligible effect [59].

Given the prevailing evidence, we opted to more directly integrate the stories and photographs of the second video with the data of the first video. To create this combination video, 50% of the content was data visualization and 50% was individuals’ stories.

In all, we constructed three video designs: data narrative, human narrative, and a mixed narrative combination of both. All three videos were matched for time and key topic coverage (the three main topics identified). More information about the videos can be found in the supplemental materials. With the hopes that the ‘best of both worlds’ approach would speak to the widest range of viewers, we anticipated that it would result in the best ‘outcomes.’

4.3 Selecting assessments

In selecting the assessments for this project, we considered the pros and cons of methods that met our criteria (see Section 3).

4.3.1 LO-A

We start with selecting an assessment for LO-A: *The viewer will observe the humanitarian crisis in Somalia*. This objective falls at the most basic level of the taxonomy (Observe/Appraisal). Based on our guidance, we opt for a standard cognitive assessment. In this case, we ask the viewer to recall details about the humanitarian crisis in Somalia. This is a basic question that will allow us to assess if the viewer has perceived the basis of the affective topic.

4.3.2 LO-B

For LO-B (*the viewer will believe that the crisis in Somalia is important*) the LO verb, *believe*, falls under the category of *strengthen verbs*⁵. The ‘noun’ part, *the crisis in Somalia is important*, could be considered an *attitude*. At this level of the taxonomy, the options for assessments are straightforward. We could ask a simple self-report question, e.g. “*How important is the current socio-political situation in Somalia?*” A more complex truth-inducing approach is likely unnecessary⁶, as we have no reason to believe the topic is sensitive or that participants might lie.

For our assessment, we can opt for a simple approach and simply ask: “How important is the current socio-political situation in Somalia?” For the scale, someone’s attitude could fall on the spectrum from thinking it is not at all important to thinking it is extremely important. To get more granularity on where someone might fall in this range, we use a Likert scale as our multiple-choice options: Not at all important, slightly important, moderately important, very important, and extremely important.

4.3.3 LO-C and LO-D

LO-C (*the viewer will believe they personally are able to make a difference in helping people in Somalia*.) and LO-D (*the viewer will believe they have a moral responsibility to help people in Somalia*.) both have *strengthen verbs* and *attitude nouns*. For assessment, we could use a self-report with scales or proxy engagement measures. Because each of our video designs is the same length, measuring a proxy engagement measure such as time spent viewing would not be an effective measure. Instead, we opt for the self-report question with a Likert scale.

One option for attitudes in particular is to modify an existing validated scale of attitudes. For example, Liem et al. (2020) [48] used the questions from the European Social Survey [26] to measure values towards immigration. In our case, we can use a validated question from a related research study [77]. By using an assessment that has already been developed, we can have more confidence that it will be actionable, by being sensitive enough to show differences between different designs and has high validity.

We asked participants to rate the following sentences.

⁵In this case, the Strength LO also measures Position because it is hierarchical. That is, if we assess the strength of the viewer’s belief, we will also be able to tell what their Position is: do they agree that the crisis in Somalia is important?

⁶For example, we could also ask *How important will other people say the situation is?* and then apply a Bayesian Truth Serum analysis [66]

Learning Objectives	Assessments	Video Designs		
		Data	Human	Combo
LO-A: The viewer will observe the humanitarian crisis in Somalia.	Q. Which of the following are topics covered in the video you just watched? Choose all that apply. A. Drought, Floods, Food Insecurity	80.6%	78.7%	77.5%
LO-B: The viewer will believe that the crisis in Somalia is important.	Q. How important is the current socio-political situation in Somalia? A. Not at all important (1); Slightly important (2); Moderately important (3); Very important (4); Extremely important (5)	3.90	3.96	3.86
LO-C: The viewer will believe they, personally, are able to make a difference in helping people in Somalia.	Q. I can make a difference in helping people in Somalia. A. Not at all (1); Slightly (2); Somewhat (3); Moderately (4); Very much (5)	2.47	2.55	2.58
LO-D: The viewer will believe they have a moral responsibility to help people in Somalia.	Q. I have a moral responsibility to help people in Somalia. A. Not at all (1); Slightly (2); Somewhat (3); Moderately (4); Very much (5)	2.92	3.01	2.90
LO-E: The viewer will practice social responsibility and donate money to the Somalia Humanitarian Response Plan.	Q. We are offering the option to donate none, part, or all of your \$2 bonus to the Humanitarian Response Plan, where the money will go to relief efforts in Somalia. If you choose to donate all or part of your bonus, we (the researchers) will donate this amount to the Humanitarian Response Plan on your behalf. How much of your bonus would you like to donate? Please enter a number between 0.00 and 2.00.	0.87	1.03	0.76

Fig. 5: Our case study evaluated five affective learning objectives with five corresponding assessments for three video designs.

- I have a moral responsibility to help people in Somalia.
- I can make a difference in helping people in Somalia.

Because we are evaluating a Strengthen learning objective, we can use a Likert response scale for our multiple-choice options: Not at all; Slightly; Somewhat; Moderately; Very much.

4.3.4 LO-E

LO-E, *the viewer will practice social responsibility and donate money to the Somalia Humanitarian Response Plan*, is the most challenging objective to assess. This learning objective falls under the category of behavior verbs and a value noun. The goal is to get the viewer to change their behavior related to a value, in this case, *social responsibility*. Specifically, we want the viewer to take the action of donating money to the Somalia Humanitarian Response Plan.

There are several ways to assess a behavior objective. Ideally, we might measure the direct behavior: the amount of money donated. However, if donation rates are low, we may need many participants. An alternative is to integrate donations directly into the protocol. For example, previous research gave participants 5 dollars as participation compensation and then gave them the option to donate some portion to a charity [76, 77]. This approach is attractive as it is incentive compatible. However, this can be more difficult to implement by increasing the cost of data collection.

An alternative is to measure a proxy for the behavior, such as a self-report of behavior intentions. For example, we can create hypothetical donation allocation questions: either telling participants that ‘someone’ was willing to make a \$10 donation in their name [11] or that they were to allocate a charity’s hypothetical \$100 to different groups of migrants [59]. A hybrid between actual donation behaviors and donation intentions is to enter participants in a chance of winning an additional \$100 and asking them what amount of the bonus they would donate if they won [50]. However, past research has indicated a bias in self-reports of donations or likely donations [7, 42], and concluded that there is a difference between hypothetical and real donations [51].

For our case study, we decided to measure the direct behavior of *actual* donations instead of a proxy behavior intention. One reason for this is that we wanted our assessment to have high sensitivity, and the direct behavior measurement is more likely to show differences between the video designs compared to if we used a self-report question. In our case, we concluded that real donations were worth the cost. At the end of our survey, we give our participants the opportunity to donate none, part or all of a granted \$2 bonus to Somalia. Participants were able to donate an amount between \$0.00 and \$2.00. We note that we were limited in the \$2 bonus amount, as the Prolific system requires us to pay the participants for the survey to approve their work, but was

not restrictive in whether or not we distributed the bonus. The bonus incentive is real money that the participant would receive if they chose not to donate.

One experimental consideration is when to run assessments in relation to the video. A pre- and post-test has the advantage of identifying both absolute (A and B are better than baseline) and relative (A is better than B) ‘learning gains.’ Focusing only on post-tests enables relative comparisons and is simpler to execute. In our case, we can largely use the post-test-only approach. However, to demonstrate absolute gains for LO-B, we employed a simplified pre-test survey.

4.4 Running the Assessment

As a demonstration of the design, we implemented our assessment via a survey on Prolific. We used the Prolific platform because it would be quick and scalable, giving us timely feedback about our designs that could be integrated into our design process. If we actually working with a partner such as UN OCHA, we could implement an A-B test on their website, showing real viewers different versions of our content. If we wanted to evaluate our designs before they are published, a crowdsourced study will be effective. Participants in our assessment earned a base compensation of \$2.80 with a bonus of \$2.00 for a total of \$4.80. On average, the survey took about 20 minutes to complete⁷.

4.4.1 Results

We analyzed data from 502 participants who participated in our survey, roughly split into three groups (173 data narrative, 164 human narrative, 165 in mixed narrative). For a summary of our assessment questions and results, see Figure 5. Naturally, running a study at this scale may not be strictly necessary for most design activities. Even ‘suggestive’ evidence from a far smaller study (on the size of a pilot) may be sufficiently actionable for a designer. We briefly describe the results and their implications for the designer’s choice. While the results are interesting, we emphasize that our focus is on using this as a case study to construct, execute and interpret affective assessments. They are specific to our three narrowly defined design alternatives.

First, we assessed whether participants observed the factors of the humanitarian crisis (LO-A). The data video design performed the best, with participants scoring at 80.6% accuracy on this assessment. The human video design (78.7%) and the combination video design (77.5%) were not far behind. Overall, the scores for the first assessment were similar across all three video designs.

⁷In addition to the questions described above, we also measured several other constructs for a different research question. These are not analyzed in this paper.

Judgments of importance were similar across all three designs (LO-B). On a scale of 1 (Not at all important) to 5 (Extremely important), the average importance ratings after seeing the video were 3.90 for the data video design, 3.86 for the combination video design and 3.96 for the people video design (moving from a pre-exposure baseline of 2.8). Thus, we see absolute learning gains but little relative differences. For LO-C and LO-D, participants were asked to rate agreement with the corresponding attitudinal statements (on a scale of 1 to 5). In general, the participants felt moral responsibility ($M=2.9$) slightly more than efficacy ($M = 2.5$), but none differed significantly by video design.

These first three objectives indicate that the LOs are achieved at roughly the same levels across all designs. However, for what is probably the most important objective (LO-E), the assessment shows there is a difference. The human video design elicited the most donations, with an average donation of 1.03 dollars. The combination video design garnered the least donations, with an average of 0.76 dollars. The data video performed somewhere in the middle (average of 0.87 dollars).

As designers, what are we to conclude from this? For our first four learning objectives, all three designs resulted in similar results. Naturally, it is possible that our assessments were not sensitive enough to measure the differences between our conditions. If we were to re-evaluate our designs, we could reassess measures from Figure 3 and opt for different assessments (e.g., a truth-inducing approach).

However, we see more variation between designs for the final assessment, donations. These results indicate that human-driven video design performs the best. Our ‘best of both worlds’ video design (showing both data visualizations and personal narratives) actually resulted in the lowest average donation. The increase from \$0.76 to \$1.03 might not seem like much in the context of the survey. However, recall that the maximum donation amount was capped at \$2.00. A 36% increase in donations could result in a meaningful amount when scaled up to an international organization like UN OCHA.

However, this result does not mean that the other videos are not useful. In this case study, we have exclusively targeted affective objectives. In reality, many communicative forms have multiple uses (e.g., to solicit donations from the public *and* inform a board of directors). The design might call for both affective *and* cognitive objectives (e.g., *recall the number of deaths* or *recall the number of clients served by the NGO*). In this case, an entirely photo-driven video would not work. The designer might opt for the data-driven video in that case. Although resulting in fewer donations (the human-only form), it would support the cognitive objectives.

Executing the assessment allows the designer to confirm (or refute) their hypothesis. It also provides useful data in the context of wicked design problems that enabled decision making. In addition, the designer can begin to develop some intuition on what works in this specific case (rather than relying on research studies for other contexts). In our case, we now have some evidence that data visualizations might numb an emotional response compared to stories and photos of individuals in need—suggesting the presence of compassion fade or statistical numbing in data visualization. While not definitive, the evidence might go against designer intuition and ultimately lead to better designs. The experience might also force thinking through alternative strategies beyond video (e.g., interactivity, scrollytelling, personalization, and other dynamic design choices to engage their audience and be more persuasive) or visualization approaches (e.g., anthropographic [58]).

5 GENERAL DISCUSSION

Past work has demonstrated that working through the creation of learning objectives and assessments will often yield better visualization design decisions [44]. However, producing good objectives and assessments in the affective domain is challenging. The work presented here provides some guidance on how a designer can utilize this framework.

One take on assessment—perhaps a controversial one—is having a good experimental design is desirable, but may not be critical for visualization *designer* decision making. For example, it would be ideal that the sampled population gave us enough power and ensured the detection of statistically significant differences under any effect size. We also would rather not select a visualization design that is much

worse than another because of some experimental design error. Some care is naturally warranted. However, the reality is that if the effect size is that small, it will make very little difference to the designer. At worst, an observed difference will be statistically meaningless and one design will be no worse than another. A corollary to this point is that not all affective visualizations need to perform complex assessments with external viewers. However, we believe that even reasoning through how a visualization *might be assessed* would be helpful to a designer for selecting one design over another in a more principled manner.

Despite our advocacy, it is worth acknowledging that even the best assessment possible—built to evaluate the most well articulated objective through the very best visualization design—may still not demonstrate a measurable change in belief or attitude. Belief instability, pliability, or centrality [19, 69] may all impact whether a change can be made by a visualization, or any kind of communication. That said, there are many cases where long term exposure to *many* forms of communication—including visualizations—can lead to real changes. A set of tools that can be applied to singular interventions or larger programs is valuable.

Finally, we emphasize that the focus in this paper is on defining objectives and assessment tools for designers when they are *in the act of building* communicative visualization. However, this does not mean that the suggested assessments will not work in other scenarios (e.g., in critique or academic research). However, these scenarios may have different goals leading to additional assessment options. For example, the proposed methods may be effective in academic studies of affective visualizations. However, in this context, the research questions may focus more significantly on mechanisms and generalizations, so the objectives will focus on smaller or different constructs. The suggested assessments are also likely to work in larger design projects (e.g., creating a full advocacy website with some visualization elements), or in long term iterative design of visualizations after deployment. More ‘expensive’ or time-consuming assessments (e.g., deeper interviews, instrumentation, etc.) may provide additional insights. It is also possible that the approaches we recommend may work in other communicative scenarios (beyond visualizations). However, as we observe in the case study, the specific features of visualizations—for example, the rhetorical ‘competition’ between pathos and logos and combinations of cognitive and affective objectives—may not be present in other formats (e.g., a simple advertising banner). The uniqueness of communicative visualizations inevitably requires a unique approach to design.

6 CONCLUSION

Learning objectives, both cognitive and affective, can help data visualization designers specify their goals: What do they expect from their viewers after engaging with a visualization? Unlike cognitive objectives, affective objectives are harder to specify and assess. That is, determining whether the viewer learned a specific data fact from the visualization is easier to determine than whether they have changed their belief on some topic. In this work, we contribute guidelines for considerations when choosing assessments when designing visualizations. We apply these guidelines in evaluating assessments from many domains (economics, advocacy, marketing, psychology). We conclude with a case study that demonstrates the creation of affective objectives (at different levels), the construction of assessments, and examples of their implementation. Assessing affective goals, while difficult, is not an impossible task. Although we do not expect designers to create perfect scientific experiments, we hope that this work enables them to create ‘good enough’ assessments. Rather than relying on intuition on what works, which can often be wrong when affective objectives are concerned, this structure provides a mechanism to build designs that are objectively more effective.

ACKNOWLEDGMENTS

Thanks to Alison Siegler, Francis Gagnon, Gabrielle Merite, Dino Citraro, and Peiyang Loh for their permission to reuse their images. The authors thank the NSF for their support of this work through NSF IIS-1815760. This work was supported in part by funding from the Rackham Graduate School at the University of Michigan.

REFERENCES

- [1] A. Abadie. Semiparametric difference-in-differences estimators. *The review of economic studies*, 72(1):1–19, 2005. 3
- [2] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Trans. on Visualization and Computer Graphics*, 27(2):946–956, 2021. doi: 10.1109/TVCG.2020.3030375 1
- [3] L. Anderson, D. Krathwohl, and B. Bloom. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives (Complete Edition)*. Longman, 2001. 1, 3
- [4] L. W. Anderson and S. F. Bourke. *Assessing affective characteristics in the schools*. Routledge, 2013. 3
- [5] A. Baillon. Bayesian markets to elicit private information. *Proc. of the National Academy of Sciences*, 114(30):7958–7962, 2017. 5
- [6] P. D. Bamidis. Affective learning: principles, technologies, practice. In *International Conf. on Brain Function Assessment in Learning*, pp. 1–13. Springer, 2017. 3
- [7] R. Bekkers and P. Wiepking. Accuracy of self-reports on donations to charitable organizations. *Quality & Quantity*, 45:1369–1383, 2011. 8
- [8] D. K. Berlo. *The process of communication: an introduction to theory and practice*. Holt, Rinehart and Winston, 1960. 6
- [9] M. Bogre. *Photography as activism: images for social change*. Routledge, 2012. 2
- [10] A. Bonica. Are donation-based measures of ideology valid predictors of individual-level policy preferences? *The J. of Politics*, 81(1):327–333, 2019. 6
- [11] J. Boy, A. V. Pandey, and J. Emerson. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*, p. 13, 2017. 7, 8
- [12] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *J. of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994. 5
- [13] R. Buchanan. Wicked problems in design thinking. *Design issues*, 8(2):5–21, 1992. 3
- [14] S. Campbell and D. Offenhuber. Feeling numbers: The emotional impact of proximity techniques in visualization. *Information Design Journal*, 25(1):71–86, Dec 2019. doi: 10.1075/idx.25.1.06cam 7
- [15] C. M. Chapman, M. Lizzio-Wilson, Z. Mirajafi, B. M. Masser, and W. R. Louis. Rage donations and mobilization: Understanding the effects of advocacy on collective giving responses. *British J. of Social Psychology*, 61(3):882–906, 2022. 6
- [16] G. Charness, U. Gneezy, and V. Rasocha. Experimental methods: Eliciting beliefs. *J. of Economic Behavior & Organization*, 189:234–256, 2021. 5
- [17] K. Coaley. An introduction to psychological assessment and psychometrics. 2014. 5
- [18] D. A. Cook and A. R. Artino Jr. Motivation to learn: an overview of contemporary theories. *Medical education*, 50(10):997–1014, 2016. 3
- [19] W. D. Crano and R. Prislin. *Attitudes and attitude change*. Psychology Press, 2011. 9
- [20] J. Cvitanic, D. Prelec, B. Riley, and B. Tereick. Honesty via choice-matching. *American Economic Rev.: Insights*, 1(2):179–192, 2019. 5
- [21] G. Daily. Somalia is on the brink of famine. *The Economist*, Dec. 2022. 7
- [22] J. Dang, K. M. King, and M. Inzlicht. Why are self-report and behavioral measures weakly correlated? *Trends in cognitive sciences*, 24(4):267–269, 2020. 6
- [23] N. Demir. Alan kurdi, Dec. 2017. Page Version ID: 813344027. 7
- [24] A. H. Eagly and S. Chaiken. *The psychology of attitudes*. Harcourt brace Jovanovich college publishers, 1993. 5, 6
- [25] P. Ekkekakis. *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge University Press, 2013. 6
- [26] <https://www.europeansocialsurvey.org/>. 7
- [27] J. Falkheimer and M. Heide. *Research handbook on strategic communication*. Edward Elgar Publishing, 2022. 6
- [28] M. Fishbein and I. Ajzen. Belief, attitude, intention, and behavior: An introduction to theory and research. *MA: Addison-Wesley*, 1975. 6
- [29] M. I. Ganesh and G. Sobliye. Communicating gender: The challenges of visualizing information for advocacy. In *Diversity and Design*, pp. 137–151. Routledge, 2015. 2
- [30] A. S. Garrett and R. J. Maddock. Time course of the subjective emotional response to aversive pictures: relevance to fmri studies. *Psychiatry Research: Neuroimaging*, 108(1):39–48, 2001. doi: 10.1016/S0925-4927(01)00110-X 6
- [31] R. Guan, E. Lee-Robbins, X. Wang, and E. Adar. Visquestions: Constructing evaluations for communicative visualizations. In *Proc. of the Extended Abstracts of the CHI Conf. on Human Factors in Computing Systems*, CHI EA '25, art. no. 575, 9 pages. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706599.3719832 1
- [32] T. M. Haladyna. *Developing and validating multiple-choice test items*. Routledge, 2004. 3, 4
- [33] K. R. Hammond. Measuring attitudes by error-choice: an indirect method. *The J. of Abnormal and Social Psychology*, 43(1):38, 1948. 5
- [34] D. G. Horvitz, B. G. Greenberg, and J. R. Abernathy. Randomized response: A data-gathering device for sensitive questions. *International Statistical Review*, pp. 181–196, 1976. 4
- [35] J. D. Jensen, A. J. King, and N. Carcioppolo. Driving toward a goal and the goal-gradient hypothesis: the impact of goal proximity on compliance rate, donation size, and fatigue. *Journal of Applied Social Psychology*, 43(9):1881–1895, 2013. doi: 10.1111/jasp.12152 7
- [36] N. Kaya and H. H. Epps. Relationship between color and emotion: A study of college students. *College Student J.*, 38(3):396–405, 2004. 6
- [37] T. Kogut and I. Ritov. The “identified victim” effect: an identified group, or just a single individual? *J. of Behavioral Decision Making*, 18(3):157–167, Jul 2005. doi: 10.1002/bdm.492 7
- [38] Kontinentalist. Abandoned at sea: The desperate journeys of rohingya refugees, Dec 2020. 1
- [39] D. R. Krathwohl, B. S. Bloom, and B. B. Masia. *Taxonomy of educational objectives: The classification of educational goals; Handbook II: Affective domain*. David McKay Company, Incorporated, 1956. 2, 3
- [40] J. A. Krosnick, C. M. Judd, and B. Wittenbrink. The measurement of attitudes. In *The handbook of attitudes, volume 1: Basic principles*, pp. 45–105. Routledge, 2018. 5
- [41] X. Lan, Y. Wu, and N. Cao. Affective visualization design: Leveraging the emotional impact of data. *IEEE Trans. on Visualization and Computer Graphics*, 2023. 2
- [42] Z. Lee and A. Sargeant. Dealing with social desirability bias: An application to charitable giving. *European J. of Marketing*, 45(5):703–719, 2011. 8
- [43] E. Lee-Robbins and E. Adar. Affective learning objectives for communicative visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 29(1):1–11, 2022. 1, 2, 3
- [44] E. Lee-Robbins, S. He, and E. Adar. Learning objectives, insights, and assessments: How specification formats impact design. *IEEE Trans. on Visualization and Computer Graphics*, 28(1):676–685, 2022. doi: 10.1109/TVCG.2021.3114811 1, 9
- [45] N. Lemon. Attitudes and their measurement. 1973. 4
- [46] E. S. E. Lewis. *Financial advertising for commercial and savings banks, trust, title insurance, and safe deposit companies, investment houses*. Levey bros. & company, 1908. 6
- [47] S. Lewis. Where are the photos of people dying of covid?, May 2020. 7
- [48] J. Liem, C. Perin, and J. Wood. Structure, empathy, and attitude. *Eurographics Conf. on Visualization (EuroVis) 2020*, p. 13, 2020. 7
- [49] R. Lifton. Beyond psychic numbing: A call to awareness. *American J. of Orthopsychiatry*, 52(4):619–629, 1982. 7
- [50] M. Lindauer, M. Mayorga, J. Greene, P. Slovic, D. Västfjäll, and P. Singer. Comparing the effect of rational and emotional appeals on donation behavior. *Judgment and Decision making*, 15(3):413–420, 2020. 8
- [51] M. Maier, Y. C. Wong, and G. Feldman. Revisiting and rethinking the identifiable victim effect: Replication and extension of small, loewenstein, and slovic (2007). *Collabra: Psychology*, 9(1), 2023. 8
- [52] D. B. Markant, M. Rogha, A. Karduni, R. Wesslen, and W. Dou. Can data visualizations change minds? identifying mechanisms of elaborative thinking and persuasion. In *2022 IEEE workshop on visualization for social good (VIS4Good)*, pp. 1–5. IEEE, 2022. 7
- [53] B. L. Martin. A checklist for designing instruction in the affective domain. *Educational Technology*, 29(8):7–15, 1989. 3
- [54] B. L. Martin and L. J. Briggs. *The affective and cognitive domains: Integration for instruction and research*. Educational Technology, 1986. 3
- [55] D. B. McCoach, R. K. Gable, and J. P. Madura. *Instrument development in the affective domain: School and corporate applications*, 3rd ed. Instrument development in the affective domain: School and corporate applications, 3rd ed. Springer Science + Business Media, New York, NY, US, 2013. doi: 10.1007/978-1-4614-7135-6 5
- [56] H. L. Meiselman, ed. *Emotion measurement*. Woodhead Publishing, 2 ed., Apr. 2021. 6
- [57] G. Merite. Climate change knows no borders. https://www.instagram.com/p/CQ_uSGCh66N/, Jul 2021. 1

- [58] L. Morais, Y. Jansen, N. Andrade, and P. Dragicevic. Showing data about people: A design space of anthropographics. *IEEE Trans. on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: [10.1109/TVCG.2020.3023013](https://doi.org/10.1109/TVCG.2020.3023013) 2, 7, 9
- [59] L. Morais, Y. Jansen, N. Andrade, and P. Dragicevic. Can anthropographics promote prosociality? a review and large-sample study. *Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems*, p. 19, 2021. 7, 8
- [60] L. Morais, D. Sousa, and N. Andrade. Evaluating a situated and physical anthropographic: An in-the-wild study. *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems*, 2020. doi: [10.13140/RG.2.2.10719.69283](https://doi.org/10.13140/RG.2.2.10719.69283) 7
- [61] D. J. O’keefe. *Persuasion: Theory and research*. Sage Pubs., 2015. 6
- [62] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini. The persuasive power of data visualization. *IEEE Trans. on Visualization and Computer Graphics*, 20(12):2211–2220, Dec 2014. doi: [10.1109/TVCG.2014.2346419](https://doi.org/10.1109/TVCG.2014.2346419) 2, 7
- [63] Periscope. U.s. gun deaths, 2013. 1
- [64] R. E. Petty and J. T. Cacioppo. *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Science & Business Media, 2012. 6
- [65] V. T. Phan. Affective learning objectives in online courses. In *Proc. of the 3rd International Conf. on Education and Multimedia Technology*, pp. 33–36, 2019. 5
- [66] D. Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004. 5, 7
- [67] K. Rall, M. L. Satterthwaite, A. V. Pandey, J. Emerson, J. Boy, O. Nov et al. Data visualization for human rights advocacy. *J. of Human Rights Practice*, 8(2):171–197, 2016. 2, 7
- [68] J. Reisman, A. Gienapp, and S. Stachowiak. A guide to measuring advocacy and policy. *Annie E. Casey Foundation*, 2007. 5, 6
- [69] M. Rokeach. *Beliefs, attitudes, and values: a theory of organization and change*. Jossey-Bass, 1968. 2, 9
- [70] L. Rost. A data point walks into a bar: Designing data for empathy. OpenVis Conference, 2017. 7
- [71] D. Rugg. Experiments in wording questions: Ii. *Public opinion quarterly*, 5(1):91, 1941. 5
- [72] J. Saks, J. L. Compton, A. Hopkins, and K. El Damanhoury. Dialed in: Continuous response measures in televised political debates and their effect on viewers. *J. of Broadcasting & Electronic Media*, 60(2):231–247, 2016. 5
- [73] D. O. Sears and C. L. Funk. Evidence of the long-term persistence of adults’ political predispositions. *The Journal of Politics*, 61(1):1–28, 1999. 6
- [74] Slovic and Slovic. *Numbers and nerves: information, emotion, and meaning in a world of data*. Oregon State University Press, 2015. 1
- [75] P. Slovic. Psychic numbing and genocide: (718332007-003), 2007. doi: [10.1037/e718332007-003](https://doi.org/10.1037/e718332007-003) 7
- [76] D. A. Small and G. Loewenstein. Helping a victim or helping the victim: Altruism and identifiability. *J. of Risk and uncertainty*, 26:5–16, 2003. 7, 8
- [77] D. A. Small, G. Loewenstein, and P. Slovic. Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2):143–153, Mar 2007. doi: [10.1016/j.obhdp.2006.01.005](https://doi.org/10.1016/j.obhdp.2006.01.005) 7, 8
- [78] P. L. Smith and T. J. Ragan. *Instructional design*. John Wiley & Sons, 2004. 5
- [79] Tactile Technology Collective. *Visualizing Information for Advocacy*. 2013. 2
- [80] M. M. Turner, R. B. Wade, R. Heo, Q. Ye, and Y. Jang. Development and validation of the commitment to social activism scale using the thurstone scaling procedure. *Communication Research Reports*, 41(3):178–192, 2024. 5
- [81] United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). Somalia: Hope fades as famine looms, Nov. 2022. 2, 6
- [82] United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). Humanitarian response plan somalia, Feb. 2023. 2, 6
- [83] D. Västfjäll, P. Slovic, and M. Mayorga. Pseudoinefficacy: negative feelings from children who cannot be helped reduce warm glow for children who can be helped. *Frontiers in psychology*, 6:134125, 2015. 7
- [84] Voilà. Freeing the us from its culture of detention. <https://chezvoila.com/project/freedom-denied/>. 1
- [85] D. Västfjäll, P. Slovic, M. Mayorga, and E. Peters. Compassion fade: Affect and charity are greatest for a single child in need. *PLoS ONE*, 9(6):e100115, Jun 2014. doi: [10.1371/journal.pone.0100115](https://doi.org/10.1371/journal.pone.0100115) 7
- [86] S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *J. of the American statistical association*, 60(309):63–69, 1965. 4
- [87] E. J. Webb. *Unobtrusive measures*, vol. 2. Sage, 2000. 5, 6
- [88] C. H. Weiss et al. Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New approaches to evaluating community initiatives: Concepts, methods, and contexts*, 1:65–92, 1995. 5
- [89] W. D. Wells. *Measuring advertising effectiveness*. Psychology Press, 2014. 6
- [90] A. Zeeberg. How images trigger empathy, Jan 2016. 7
- [91] P. G. Zimbardo and M. R. Leippe. *The psychology of attitude change and social influence*. Mcgraw-Hill Book Company, 1991. 6