Are We Healthier Together? Two Strategies for Supporting Macronutrient Assessment Skills and How the Crowd Can Help (or Not)

SARAH M. HARMON, Bowdoin College, USA
ELIZABETH M. HEITKEMPER, The University of Texas at Austin, USA
LENA MAMYKINA, Columbia University, USA
MARIA L. HWANG, Fashion Institute of Technology, USA

Learning macronutrient assessment skills can support improved health outcomes and overall wellbeing. We conducted two Mechanical Turk studies to investigate how users might benefit from the crowd's input in macronutrient assessment education. We first determined whether the wisdom of the crowd alone would provide users with enough insight to arrive at accurate macronutrient estimates. Next, we tested two methods of teaching macronutrient assessment skills (Comparison and Decomposition) and analyzed their effectiveness. Results from these studies indicate that while the crowd alone may not be sufficient to support this type of education, users may yet benefit from access to community-generated photos and labels while they use either the Comparison or Decomposition strategy.

CCS Concepts: • **Human-centered computing → Collaborative and social computing**; *Empirical studies in collaborative and social computing*; • **Applied computing → Education**; *Collaborative learning*; **Life and medical sciences**; *Consumer health.*

Additional Key Words and Phrases: carbohydrate estimation, nutrition literacy, learning-by-example

ACM Reference Format:
Sarah M. Harmon, Elizabeth M. Heitkemper, Lena Mamykina, and Maria L. Hwang. 2022. Are We Healthier Together? Two Strategies for Supporting Macronutrient Assessment Skills and How the Crowd Can Help (or Not). *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 492 (November 2022), 25 pages. https://doi.org/10.1145/3555593

1 INTRODUCTION

Unhealthy eating contributes to the global obesity epidemic, which was declared a major health problem by the World Health Organization in 1997. Nearly 10% of the world's population was obese in 2015 and studies indicate obesity numbers have tripled between 1975 and 2016, further pointing to the growth of this problem [19, 38]. Additionally, data from the Centers for Disease Control and Prevention in 2019 shows there are now 12 states in the US with adult obesity prevalence at or above 35%, which is up from nine states in 2018. Furthermore, this is projected to increase with the impact of COVID-19 [62]. Prior research indicates that a significant reason why people make unhealthy food decisions is low nutritional literacy with particular emphasis on people's inability to accurately estimate the nutritional composition of a meal [46].

Authors' addresses: Sarah M. Harmon, Bowdoin College, Brunswick, Maine, USA; Elizabeth M. Heitkemper, The University of Texas at Austin, Austin, Texas, USA; Lena Mamykina, Columbia University, New York, New York, USA; Maria L. Hwang, Fashion Institute of Technology, New York, New York, USA.

2573-0142/2022/11-ART492 $15.00
https://doi.org/10.1145/3555593

Nutrition literacy and the ability to estimate the nutritional components of meals is particularly important for obesity prevention and reduction as well as the self-management of chronic conditions that frequently co-occur with obesity such as diabetes, kidney disease, and cardiovascular diseases [18, 44, 75]. Managing these conditions often requires dietary restrictions of specific nutrients, such as carbohydrates or sodium. However, in order to integrate these changes and restrictions, an individual needs to be able to identify and estimate the quantity of these nutrients in foods. The ability to estimate different food groups or macronutrients, particularly carbohydrates (sometimes referred to as "carb counting"), in a meal directly impacts glycemic values, overall weight, and one's wellbeing [72, 74].

Previous studies have shown that estimating nutrition in meals presents considerable challenges for everyone. For example, people with low literacy have difficulty interpreting nutrition fact labels [10], often leading to the miscalculation of the amount of consumed nutrients [15, 41, 70]. Furthermore, even professional dietitians find making accurate nutritional estimations difficult. Expert dietitians tend to underestimate portions, and therefore calories, of large meals [13]. Prior research has demonstrated that healthy eaters underestimate calories more than unhealthy eaters [12]. Complex meals with many components and ingredients add another layer of difficulty with nutrition assessment and further demonstrate why this is a universal struggle and not one that only groups with low literacy or poor numeracy skills face. To address this significant challenge, apps and games to support nutrition tracking and estimation have been created [4, 36, 37]; however, as the "Related Work" section will clarify, current tools have clear limitations. Solutions created to target nutrition estimation have primarily focused on the tracking of foods consumed throughout the day via two specific approaches to logging of meals and/or meal components that are: 1) utilization of nutrition databases, and 2) image analysis of submitted user-generated meal photos. Unfortunately, both of these approaches suffer from problems and unintended consequences. While large nutritional databases such as Nutritionix include a vast amount of common foods, grocery and restaurant items, research has noted a trade-off between the burden of logging meals and the accuracy and completeness of meal logs [52]. Currently, the application of user-generated meal photos has been limited by either lack of nutrition data for the photographed meal [20, 59] or the need to employ an extensive number of people to assess the meals [31].

These constraints suggest that users need more assistance with nutrition estimation [9] to minimize many of the issues in current technology-assisted solutions and to truly facilitate nutrition knowledge and healthier eating. One promising area for examination is a human-in-the-loop solution for nutrition estimation. Despite the difficulties everyone (even experts) face with accurately estimating the nutrition of a meal, prior work has proposed that the wisdom of the crowd might be applied to lessen the burden of making healthy food decisions [40]. It was thus wondered if users could make correct estimates or nutrition and meal choices with the help of one another (i.e., through access to the crowd's wisdom). Building on a prior study from 2011 by Noronha et al. [59] that used a complex system of Mechanical Turkers to estimate user's meals for them, our first study (Study 1) examined if a more lightweight use of the crowd, via a community board (CB) that presents the most commonly-selected meal choices, could assist users enough to help them arrive at the correct choice. Results from Study 1 suggested that users could *not* make the correct nutritional estimation choice *with access to the untrained crowd's insights alone*. We hypothesized next that the crowd might yet be able to provide benefits assuming users were taught a nutritional estimation strategy at the same time.

In our subsequent study (Study 2), we developed two approaches to teaching nutrition estimation, referred to in our study as Decomposition and Comparison, that were based on previously established educational methods. Conducted with Amazon Mechanical Turkers, Study 2 examined the effectiveness and experience of both approaches versus a control group. The element of crowdsourcing in this study was derived from an approach suggested by prior work in meal tracking [63]. This approach consists of crowdsourcing meal photos and then having multiple users evaluate them through participating in the educational activities, which has the potential to produce a burgeoning database of real world meal photos with corresponding nutrition assessments. With time, these meal photos might be leveraged to fill a gap in the current approaches to meal/meal component logging.

We found that user learning improved in both experimental conditions compared to the control. Additionally, despite concerns about their accuracy, users felt both educational strategies were relatively easy and useful. Overall, this work identifies several lessons learned to inform the design of future collaborative apps for macronutrient assessment education.

The contributions of these two studies include:

- A lightweight version of the crowd alone is not able to support the lay user enough in identifying the correct macronutrient assessment from a meal photograph.
- Improvement in carb estimation skills can be seen when users are taught either the Decomposition or Comparison strategies using our novel application. We used two evidence-based approaches to macronutrient estimation with crowdsourced meal photographs to inform future work addressing limitations present in previous solutions.
- User feedback points to the potential for crowdsourcing meal photos and labels from users trained by either strategy in supporting future macronutrient assessment education and inspiring healthier eating.

2 RELATED WORK

2.1 Scanning or Selecting Food Items From a Nutrition Database

The first class of nutrition tracking and food estimation apps relies on using the structured recording of meals by scanning barcodes from pre-packaged foods to select food items/meals from nutrition databases [2, 57, 60]. This approach is commonly employed by popular commercial apps like MyFitnessPal and has been used in many past studies [37, 42, 50, 64, 66]. Prior work by Siek et al. [77] used an electronic food intake monitoring application for patients with chronic kidney diseases to report their food intake through barcode scanning and voice recordings. They found that only 60% of the food items scanned through barcodes could be identified because the open source food database lacked variety of food items that were available at low-cost stores [75, 76]. Furthermore, nutrition databases tend to be concentrated on commercial and fast food items and lack many ethnic and homemade foods, which are often healthier than easy to track items [16, 87]. Moreover, even if databases contained a sufficient breadth of foods and information, studies have shown that selecting meals from nutrition databases often leads to inaccurate records, as individuals often select meals that only loosely resemble their own [71].

One mobile app named *Pirate Bri's Grocery Adventure* uses a nutrition database to support healthy food choices. Users are intervened on at the time of shopping for groceries, inspired by research that shows apps leveraging real-world decisions increase users' nutrition knowledge and self-efficacy [7, 8]. This gameful app has a pirate dog nutritionist present food literacy challenges, and as the users fill their carts they are prompted to visualize each item's macronutrients (e.g., carbs, proteins, etc.) using colours that highlight low, moderate, or high amount. However, the items in their "carts" are limited to only scannable items. For non-scannable food items the user needs to input the food item and its associated nutrition manually, which increases the burden of logging and requires a level of literacy in order to properly identify the non-scannable items correctly. Since barcodes are only available for pre-packaged foods, other studies have shown that an unintended consequence of using these types of apps is that individuals will opt to increase consumption of less healthy pre-packaged foods, as opposed to homemade meals, to simplify data entry [3, 22]. *Carb Counting with Lenny* is an alternative app in which users are taught carb counting via single player or online competitive games [53]. Users can try building a meal that meets a particular target value of carbohydrates, as well as guess which food item has more carbohydrates. In another mini-game, users guess whether a displayed food is an example of a carbohydrate or not. There is also a food guide available that simply lists the carbohydrate amount and serving size for each image in a library of food items. It is possible to add custom pictures of food items, but the user must provide the associated data. Apps like *Carb Counting with Lenny* and *Pirate Bri's Grocery Adventure* are a great start toward building macronutrient assessment skills, but the food and meals used as exemplars are limited in scope. For *Carb Counting with Lenny*, the app does include a means for users to submit their own food entries, but the overall database is limited and consists of mostly "typical" Western foods. This lack of diversity might not translate well to real world situations for many users.

A study by Merler et al. [54] created a large data set of "in-the-wild" food photos from crowd-sourced mechanisms and ultimately made one of the largest food databases of photos from real people. However, a limitation of this work was the focus on capturing photos of only single food items and not full meals. By not capturing full meals, this work missed an opportunity to address another commonly-identified issue with the nutrition database approach, i.e., that food item/meal component logging does not reflect the reality of how people most frequently tend to eat—complex meals that consist of numerous food items. The need to log each individual component is not only time consuming and error prone, but it also does not reflect actual eating habits. As a result, this approach often does not translate well to the nutrition estimation of full meals.

2.2 Image Analysis of User-Generated Meal Photos

An alternative set of solutions have focused on having users track their intake by uploading a photo of their meal and estimating the nutrition in their meal using the image. Prior work has utilized wearable devices such as SenseCams to automatically capture meal photos [82] and reduce the burden of taking meal photos. Similar tools analyze the meal images using a variety of methods that include using a group of experts like registered dietitians, the general public through crowdsourcing, and more recently through applying computational image analysis techniques [5, 54, 55, 68, 69].

Common approaches include using machine learning algorithms to recognize components of meals based on training datasets of images of meals labeled by experts or crowd workers. For example, [68] used a convolutional neural network (CNN) approach in conjunction with a computer vision system to improve food labeling and calorie estimation performance. This approach included segmentation of the image using computer vision and then classifying different segments according to their nutrition using CNN. This team used a labeled dataset of restaurant meals to train a CNN-based multi-classifier that associated depth of each pixel from a single RGB image to the volume of food depicted in the image [55]. The classifier was then applied to non-restaurant food images to arrive at their caloric estimation [54].

In the recently-developed *GlucoGoalie*, an app focused on providing personalized nutrition goals, users upload images of their meals that are assessed by registered dietitians for their macronutrient content. Once enough meals have been recorded and estimated, *GlucoGoalie* [27, 31] uses an expert system to generate nutritional goal recommendations for people with type 2 diabetes who are working on controlling their blood glucose (BG) values. The app uses a machine learning algorithm for detecting patterns of association between nutrition in user-uploaded meal photos and changes in BG values. An example of a personalized goal created from GlucoGoalie is: "For high carb lunches, decrease your carbs to be about 2.5 carb choices (38g). For example, 1 carb choice is 1 slice of whole wheat toast, 1/3 cup of plantains, or 1/3 cup of brown rice." These recommendations include both the basic carb counting method that uses carb choices, which are based on food exchange lists, and the more advanced method that uses grams [23] and include examples for people less familiar with using these approaches. Through the expert knowledge that is incorporated in these nutritional goals and examples, the app intends to provide the user with the knowledge to accomplish these goals and learn from them as they go through trial and error with their own everyday meals.

Since *GlucoGoalie* is based on the users' submitted meal photos, the meals are quite diverse and personally meaningful, but the app has its own set of limitations. Users have reported becoming frustrated at the need to photograph all their meals prior to eating, essentially watching their food get cold and requiring patience from meal companions. This finding is similar to results of a 2015 study on a lightweight photo-based food journal by Cordeiro et al., that showed barriers to taking meal photos were forgetting, too high a level of difficulty, and feelings of stigma when eating outside the home [20]. These barriers were further validated in a study published the subsequent year that was also conducted by Cordeiro and additional researchers [22]. Furthermore, in this approach, while personally meaningful, since the nutrition evaluation is based on their meals, users are not learning from a variety of foods and might not be well positioned when faced with foods that are not included in their regular routines.

A different area of research has used 3D shape reconstruction to identify volumes of different foods within identified segments and thereby arrive at nutritional composition of foods using a nutritional database. Anthimopoulos et al., [5] used this approach to develop a smartphone solution for helping individuals with type 1 diabetes estimate carbohydrates in their meals [69]. These approaches that rely on computer vision has important advantages, as they only require their users to photograph their meals, thus minimizing user burden associated with meal logging. However, these approaches continue to be limited in their ability to recognize a wide variety of meals with ingredients that may not be visually apparent [47, 68, 86]; limited to commercial, and/or fast food chain items or food items belonging to a predefined food category [16, 87]; or limited in focusing on outcomes on one nutritional aspect such as estimation of calories [48, 55, 56] or portion sizes only [14, 15, 58], or one particular population with limited generalizability [5, 25, 26].

2.3 Methods for Teaching Nutrition Estimation

Since many prior apps include elements of teaching nutrition estimation skills, for study 2 we sought to examine traditional strategies from the literature to guide the development of our approach to teaching nutrition estimation. The most commonly used strategy in nutrition education practice is called MyPlate [61], and is currently the method advocated by the U.S. Department of Agriculture (USDA). In the MyPlate method, people learn to identify food groups, and then are able to use the MyPlate visual (e.g., a plate of food that consists of icons representing an appropriate amount of each food group that should be eaten) to help them eat similar portions of each food group on their own plate [39]. MyPlate visually depicts how a person's meals should generally look, allowing for comparison to their actual plate of food. Prior work has found this approach is frequently used with young children and college students making it well suited for adults of all literacy and numeracy levels [43, 84, 85]. For our study, we developed a process rooted in the MyPlate method, referred to as *Comparison* from here on.

Another well established strategy for nutrition estimation was developed for the context of chronic disease management and is based on the strategy long used to assist people with diabetes (primarily type 1 diabetes) with estimating their meals and nutrition, as this is a key element of their treatment plan, called carbohydrate (carb) counting [79, 80]. In carb counting, people break down each meal into its macronutrient components (e.g., carbs, proteins, fats, etc.), with a focus on carbs, and estimate the amount of each in every meal [49]. Through this approach, people track how much of each macronutrient they consume throughout the day and are able to make adjustments to their choices based on their daily running totals [33]. This approach (referred to in this paper as the *Decomposition* approach) is much more complex and involves numerous multi-step calculations, which can be difficult [32, 51]. However, the Decomposition approach also has been used in much of the aforementioned prior work in this space. The Decomposition approach is also more concrete than the Comparison approach because it involves closely examining the individual components of the target meal. In contrast, the Comparison approach involves making an educated guess about the target meal by way of considering a similar meal. It is thus possible that the Decomposition approach could be more effective. Despite the frequent application of both strategies, there are not any side-by-side assessments to examine if any measurable advantages or disadvantages exist when applied to digital solutions. As the Decomposition strategy assesses meals in terms of macronutrients (and not food groups), we modified MyPlate's Comparison strategy to also be in terms of macronutrients to facilitate our comparison.

Prior work (such as *Carb Counting with Lenny* and *GlucoGoalie*) has already leveraged the carb counting method in different ways. In our work, we aimed to distinguish carb counting from the Comparison strategy and explore user feedback relating to features that support greater learning and engagement through a social context. While *Carb Counting with Lenny* has a social component in that users can play games against each other, this is only in the context of competition through mini-games and does not include features that allow for people to actively learn from one another. This also applies to *GlucoGoalie* and *Pirate Bri's Grocery Adventure*, since they have no features to support social interaction. The studies presented here thus present a start toward filling the research gap around the impact of the crowd on learning nutrition estimation. Finally, it is unclear that users are able to *reliably* learn from these experiences, or if there are other learning strategies that may be useful. Past work has expressed doubt that these types of apps and their approaches would result in engaged learning [45, 63], and/or in better learning than traditional educational websites and modalities [78]. Recording user feedback and assessing whether (and to what degree) the Decomposition or Comparison strategies could actually result in learning were thus both important research goals.

3 STUDY 1: (NOT) LEARNING NUTRITION ESTIMATION FROM THE CROWD

3.1 Research Questions

Our research questions for Study 1 were as follows:

(1) Does practicing nutrition estimation with access to crowd answers and reasoning (i.e., the community board) improve users' nutrition estimation skills?
(2) How frequently do users perceive the crowd answers and reasoning to influence their final choice?
(3) How does exposure to incorrect information from the crowd answers and reasoning influence users' final choice?

3.2 Methods

3.2.1 Meal Photo Dataset. Study 1 relied on an existing dataset of 73 "in-the-wild" meal photographs labeled with macronutrient and health goal suitability information [27]. The photographs were collected (initially unlabeled) by users participating in a series of meal diary studies. The photographs were annotated by a team of 20 nutrition experts who were recruited via flyers given to local university nutrition programs. All experts had to have earned the registered dietitian (RD) credential to serve as an expert.

Expert annotations included macronutrient assessments for each meal. This meant an RD expert would estimate the amount (in grams) of protein, carbohydrate, fiber, and fat in the meal based on a meal photo and an accompanying short description of the meal components. Meal photos were then grouped into sets of four that correlated to a relevant health goal. The experts also ranked these sets of meal photos in terms of their suitability for achieving this health goal. For example, if a user's health goal was to manage type 2 diabetes, the associated nutritional goal was to decrease carb consumption. For each set of four meal photos, expert RDs ranked the photos in the following way: one best choice, one second best choice, and two equally less desirable choices. Rankings were based on the content of their macronutrient assessment. Each goal had 5 unique sets of meal photos associated with it. A final check was done by the study team to assess for photo resolution quality and content clarity.

3.2.2 App Development and Onboarding. We developed a mobile (Android) app to test our hypothesis that lay users might be able to make healthier meal choices for a particular health goal if they could see what the crowd (e.g. majority) believed to be the correct choice. We followed the same recruitment process as in similar prior work studying meal and nutrition assessments by lay users [63] and made the study available on Amazon Mechanical Turk and social media sites for volunteers to sign up. Participants (30F, 38M; $N_1 = 68$) were screened with the following inclusion criteria: 1) own an Android device, and 2) be 18 years of age or older. After accepting the terms of the study found in the consent form (approved by the IRB from the researcher's university), they downloaded an APK file and installed the app on their Androids. They were then directed to (1) take a pre-test, (2) use the app, and finally (3) take a post-test and experience survey to complete the study. The pre- and post- tests helped us determine if using the app resulted in nutritional learning. All participants were paid $6 for their participation with an expectation to complete the study between 20 to 30 minutes.

3.2.3 Using the App. Through the app, users participated in five rounds in which they attempted to select a meal photo that best fit a particular health goal. There were a total of four health goals in the app, each with a brief background story as to why a particular nutritional goal should be pursued. For example, *"X has Irritable Bowel Syndrome. Their daily life is greatly influenced by the way their digestive system behaves. One part of their treatment plan is managing their diet. Their dietitian has given them the goal of* ***increasing the amount of fiber*** *in their diet."* The other three stories and goals were: training to run a marathon (associated nutritional goal: **increase carbs**); managing type 2 diabetes (goal: **decrease carbs**); and dealing with overweight (goal: **decrease fat**).

During each round, they could choose one of four meal photos that they felt best fit the goal (see Figure 1). Each meal photograph was accompanied with a text description of its contents (e.g., fat free Greek yogurt with one cup of grapes and a coffee). The meal choices shown were randomized within each round, as well as the order of the rounds themselves. Users were ultimately shown twenty meal photos overall (five rounds with four unique photos per round). Figure 1 shows an example of four meals that was associated with the nutritional goal of decreasing carbohydrates. As illustrated in Figure 1, meals that were similar in nature (e.g. [1] yogurt with fruit and coffee; [2] yogurt, bread, boiled eggs, and fruit; [3] nuts and fruit; [4] cashew nuts) were chosen for each round.

After they selected the meal photo that best fit their goal, they could view what other members of the crowd had selected as their choices through a Community Board (CB) as shown in Figure 1(a). The CB also displayed why each member of the crowd made that choice (e.g. their reasoning) as shown in Figure 1(a) and (b). After viewing the CB, they could then change (or not change) their answer based on this information. Figure 1 displays an example of what the user saw during this process. Users were informed that the CB was made of contributions from their peers (not trained experts).

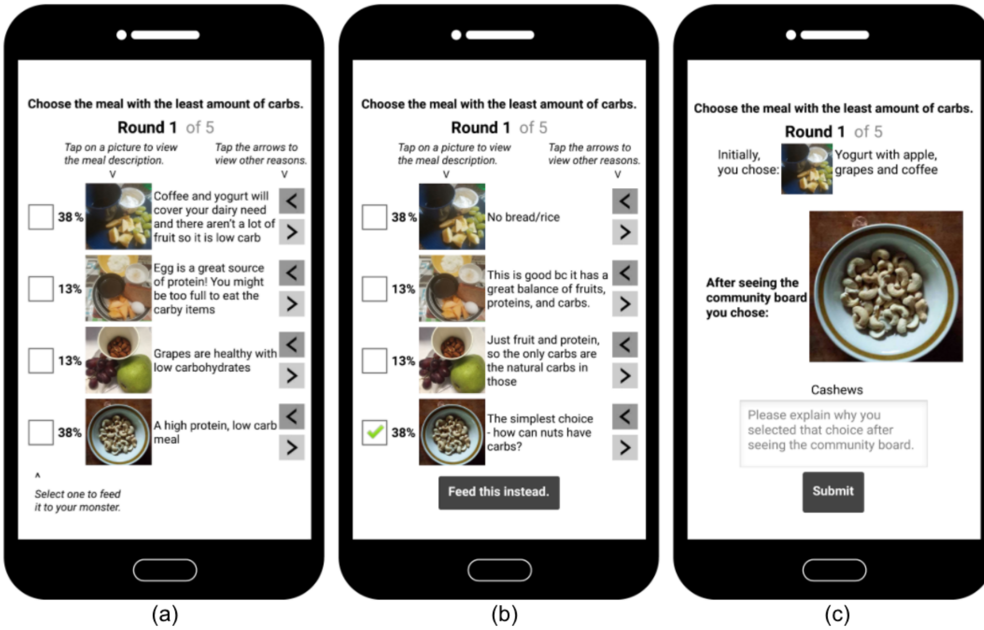

Fig. 1. An example of how the Community Board (CB) information was shown to a user in the mobile application from Study 1, which could result in that user changing their individual opinion. In every round, users could view how the crowd ranked each of four provided meal photos in fitting a particular nutritional goal (1a). Reasons for those rankings were also provided next to the meal photos and could be "scrolled" through by clicking the forward and backward buttons. After viewing the crowd's opinion, users could change their answers by clicking on a checkbox associated with a particular photo (1b) and finally confirming their answer (1c).

Since the CB was based on crowdsourced evaluations by untrained users, the CB did not guarantee correct information at all times. The CB simply represented the wisdom of the untrained crowd (or lack thereof). In two of the five rounds this issue was realized, as the crowd wisdom selected the wrong meal photo as the best choice for the goal. However, the meal photo dataset itself was provided to us with expert annotations that included ordinal ranks for how well each meal fit a particular health goal. We thus were able to determine how accurate a particular user or the overall CB was (i.e. how close their selections matched with trained experts).

3.2.4 Recall Post-Test and CB Experience Survey. We were interested to see if users learned anything from the CB experience (Research Question 1), what their perceptions were (Research Question 2), and if exposure to incorrect information from the crowd influenced users' final choice (Research Question 3). To gather this information, all users took a pre- and post-test outside of the mobile application.

The pre-test consisted of questions where users had to identify which of two meal photos better fit a specific macronutrient-focused goal (e.g. which meal was higher in carbohydrates) — similar to Figure 3. Participants did not receive feedback on the accuracy of their responses. Like the pre-test, the post-test macronutrient assessment presented users with a photograph and brief description of two meals and asked them to select which better fit a given macronutrient goal. Two questions related to the CB experience were posed to the users: 1) How often did you agree with the rest of the community? (Research Question 2) and 2) Did the community board influence your final meal choice? (Research Question 3). User responses used a scale with six choices that consisted of ‘Always;’ ‘More than half of the time;’ ‘About half of the time;’ ‘Less than half of the time;’ ‘Never;’ and ‘I did not look at the Community Board.’

3.3 Results

Sixty-eight participants’ ages ranged from 18 to 65 with a median age of 25-34 years old. Participants were evenly split across gender (44% female) and education (50% above high school/GED). Most participants reported not working professionally in nutrition (75%) and having no prior experiences with using health applications (60%). In choosing which goal to select, participants preferences were relatively evenly distributed across each of the four options: ‘Type 2 Diabetes’ (27.9%), ‘Weight-loss’ (27.9%), ‘Running a Marathon’ (27.9%), and ‘Irritable Bowel Syndrome’ (16.2%). All participants averaged around 7 minutes and 12 seconds to go from round 1 through 5 in the app (with the longest lasting 20 minutes), excluding the time taking the pre and post-tests.

No statistically significant learning (recall or transfer of learning) occurred in any of the groups from pre- to post-test. User experience with the CB also did not vary significantly across any demographic categories. As described in the study methodology, two main questions were asked to assess whether viewing the CB’s information had an impact on the users’ choices selecting a meal. In responding to the first question, “How often did you agree with the rest of the community?,” 17 out of 68 users (25.0%) selected ‘Always;’ 29 users (42.6%) chose ‘More than half of the time;’ and 13 users (19.1%) selected ‘About half the time.’ ‘Less than half the time to never’ consisted of the remaining 9 users out of 68 (13.3%).

The second question, “Did the community board influence your final meal choice?,” resulted in 16 out of 68 users (23.5%) selecting ‘Always;’ 11 users (16.2%) choosing ‘More than half of the time;’ and 16 users (23.5%) selecting ‘About half the time.’ ‘Less than half the time to never’ consisted of the remaining 25 users out of 68 (36.8%). In addition, 31 out of 68 (45.6%) users changed at least one of the five answers after reviewing the crowdsourced opinions of the CB. Among the 31 who changed their meal choices, 25 of them ended up going from an incorrect to correct answer (80.6% among those who changed answers; 36.8% out of the total 68 users). Among the 31 users who changed answers, 21 changed their answers in more than one round (38.7%). One user changed their meal choices in four out of the five rounds, which all led to a correct answer.

Among the five randomized rounds each user played, 38 participants (out of 68) were exposed to CB results that displayed the wrong meal as the “best meal” for two out of the five rounds. In other words, the meal chosen by the majority of the CB did not match the meal that the experts labeled as best fitting the nutritional goal. Among those 38 participants, 24 (63.2%) started with the wrong meal choice. After exposure to the incorrect CB information, they remained with the incorrect meal choice for either one or both of the two rounds. Ten (41.7%) went from incorrect meal to incorrect final meal choice post CB exposure for both of the two rounds. The remaining 14 (58.3%) went from incorrect meal choice to incorrect final meal choice after CB exposure for one round of the two rounds that displayed incorrect CB information.

4 STUDY 2: ASSESSING STRATEGIES FOR LEARNING CARBOHYDRATE ESTIMATION THROUGH DIGITAL INTERVENTION

4.1 Research Questions

Our research questions for Study 2 were as follows:

(1) How do each of the three strategies (Comparison vs. Decomposition vs. Control) impact nutrition estimation learning?

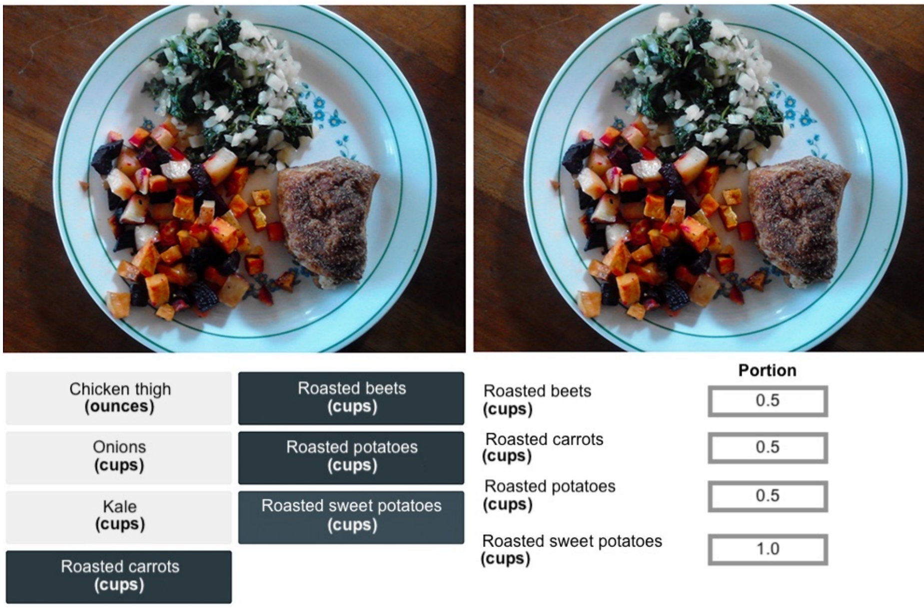

Fig. 2. Left: an example of what a Decomposition strategy user sees as they select ingredients containing carbohydrates. Right: the same user estimates portion size for the meal. Both steps provide information to help the user estimate the total carbohydrates in the meal.

(2) Which of the three strategies (Comparison vs. Decomposition vs. Control) creates a more positive user experience?

4.2 Methods

4.2.1 Study Overview. Taking lessons learned from Study 1, we deployed a between-subjects study design on Amazon Mechanical Turk to evaluate two strategies (Decomposition vs. Comparison) against a control for carbohydrate estimation from user-generated meal photos for this study (e.g. Study 2). The same dataset of annotated meal photos (from Study 1) was used for Study 2. The target population for this study was also adults 18 years old or older and recruiting was conducted through Amazon Mechanical Turk (AMT). The Decomposition (11F, 20M), Comparison (5F, 26M), and Control (12F, 19M) group had 31 participants each (N_2 = 93). Mechanical Turkers were paid $6 for their participation. This was determined by calculating the appropriate amount for the participation duration (approximately 20 to 30 minutes) based on the Department of Labor minimum wage information.

4.2.2 Two Nutrition Estimation Strategies. The *Decomposition* strategy was based on one currently utilized in diabetes education [6, 12, 13, 79], for helping people estimate the total amount of macronutrients in a meal by estimating each macronutrient individually for each meal. This approach requires individuals to mentally break their meals into each macronutrient group (e.g. "sweet potatoes are carbs"; see Figure 2 left), estimate portion sizes of these components (e.g. "2 cups"; see Figure 2 right), and then using conversion tables (e.g. "1 cup of sweet potatoes = 27g of carbohydrate") arrive at their nutritional content (e.g. "2 cups of sweet potatoes = 54g of carbohydrate"). This method, however, can require many steps, particularly for complex meals that include multiple ingredients such as a salad. As a result, this approach to nutritional estimation can be burdensome and historically has led to abandonment from users. Our approach mirrors this

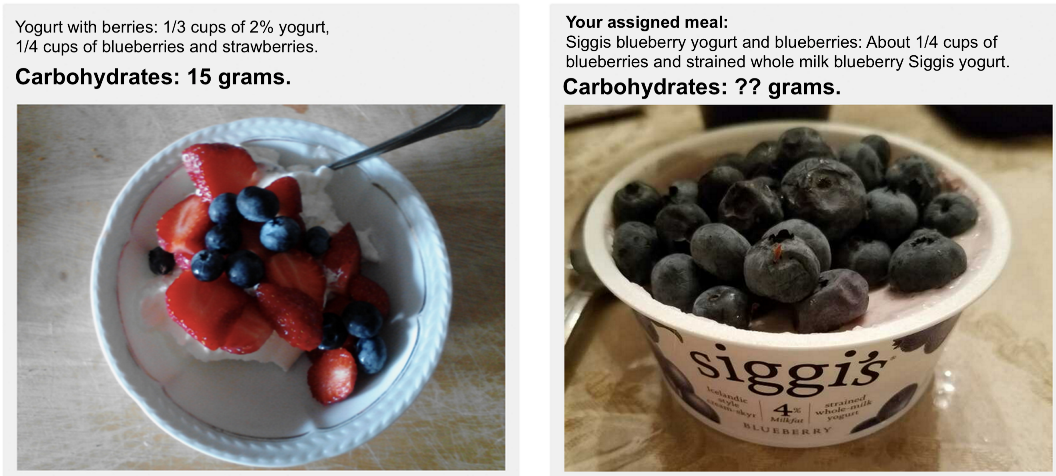

Fig. 3. A sample Comparison strategy question. At the left is an anchor meal that has quantitative (i.e., carb values) and qualitative descriptions, while the right picture is a target meal. The user must estimate the carb content of the target meal based on the meals they see on the left side of the screen. The right meal is compared to five different similar anchor "left meals" before a concrete carb estimation has to be made. During the five rounds of comparison the user only has to indicate whether the right target meal has less, more, or equal amounts of carbs compared to the left meal.

process but is scaffolded through the app structure and organization to help reduce these burdens while users learn and practice these skills.

The second strategy is based on the MyPlate method of nutrition estimation that relies on comparing your meal to a standard meal (e.g. the *Comparison Strategy*; see Figure 3). For our study, we modified this approach and had users estimate the nutritional composition (NC) of their meals by comparing five "in-the-wild" meals, not to a gold standard hypothetical plate, but to meals captured by others for which NC is known. Using this method, individuals visually search through a collection of meal photographs for which NC has been previously obtained and find meals most comparable with their target meal based on the components and ingredients, instead of mentally examining each ingredient of their meal (as one would do in the Decomposition Strategy). They can then use these comparable, crowdsourced meals as benchmarks for estimating macronutrients (e.g. carbs) in their target meal.

Since no learning was observed in Study 1, we focused on creating lightweight versions of learning strategies for Study 2. The Comparison Strategy not only mimics the idea behind MyPlate (currently advocated by the USDA), but also at its core is inspired by Study 1's basic premise of showing and utilizing full, "in-the-wild" meal photographs collected from the crowd with qualitative descriptions. In contrast, the Decomposition strategy has been previously applied most frequently in prior work and might be a more robust way of teaching nutrition estimation skills because once the process is learned and the conversion tables are memorized, the users' ability could be forever improved. The objective of the Study 2 was to compare the effectiveness of the Decomposition and Comparison strategies on individuals' ability to accurately estimate carbs of "in-the-wild" user-contributed meals. We were interested if either or both of these strategies were more effective compared to the control group, and whether one strategy led to better learning than the other.

Condition	n_1	**Activity-GT Mean**	SE_1	n_2	**(|Post-GT|-|Pre-GT|) Mean**	SE_2
Decomposition	155	15.565	21.608	186	-13.414	3.729
Comparison	154	14.229	23.512	186	-18.566	3.727
Control	155	21.161	29.213	186	-2.209	3.731

Table 1. Means and standard errors (SE) for participants' nutritional estimate accuracy for Activity (during the experimental activity), Pre-test, and Post-test meals for all conditions. GT is the ground truth, i.e., the number of carbohydrates as determined by expert dietitians. n_1 is the number of rounds/meals (five) multiplied by the number of participants in each condition (31). n_2 is the number of questions for the pre and post-tests (six) multiplied by the number of participants (31).

4.2.3 Control Group. The control group involved no strategy that assisted the user in any way. The users simply were presented with a meal photograph with ingredients listed and had to estimate the total carb amount in each meal. While the Decomposition and Comparison methods incorporated a nutrition estimation strategy that could help a user arrive at a more accurate estimate of carbs in a meal photograph, the control group lacked any strategy that could assist them in arriving at a better or more accurate nutritional estimation.

4.2.4 Study Procedure. Each condition was given the following in the same order: informed consent, introduction to survey flow and definition of carbohydrates (carbs), demographics questions, six randomized pre-test questions, experimental activity, six post-test questions, and open-ended questions. Each of the conditions had five assigned meals that were provided in the same order. Additionally, there were six randomly ordered pre- and post-test items for all conditions. We asked two open-ended questions at the end of the survey: 1) What do you think of this way of assessing nutrition in meals? and 2) Discuss your thoughts on seeing meals prepared and posted by other people?

4.2.5 Data Analysis. Statistical Package for the Social Sciences (SPSS version 28, licensed 2021) was used to analyze the data and alpha was set at .05. Three interdisciplinary members of the research team used inductive thematic analysis to analyze the open-ended feedback we received. First researchers met for a 1-hour collaborative coding session during which they read the transcripts together, discussed the meaning and significance of the different segments, and developed an initial coding scheme. Then investigators combined categories into higher level themes.

4.3 Results

4.3.1 A Priori. A chi-square test of independence showed no statistical significance in gender $\chi^2(2)=4.395$, $p=.111$; age $\chi^2(8)=10.479$, $p=.233$; prior nutrition knowledge $\chi^2(8)=9.850$, $p=.276$; and computer usage $\chi^2(6)=7.028$, $p=.318$; but there was a statistical significant difference with education level $\chi^2(4)=10.197$, $p=.037$ among the Comparison ($n=31$), Decomposition ($n=31$), and Control ($n=31$) conditions. An ANCOVA test revealed no statistically significant differences when assessing if education level affected the pre-test ($F(2, 555)=1.617$, $p=.199$).

4.3.2 Carb Estimations: Decomposition vs. Comparison vs. Control. For all conditions, every user estimated the carb content of five assigned, non-evaluated meals (during the experimental activity) using their randomly assigned method. To measure participants' accuracy, we subtracted the ground truth (GT) nutritional estimate for each meal (GT is the amount of carbohydrates assessed by expert dietitians) from the participant's nutrition estimation of that meal. When this value is close to zero, it means there is little difference between the participant and expert's assessment (i.e., the desired outcome). The greater this value is, the greater the difference between the participant and expert's

I group	J group	Mean Difference (I-J)	SE	p
Decomposition	Control	-11.205	5.279	0.034*
Comparison	Control	-16.357	5.275	0.002**
Decomposition	Comparison	-5.152	5.272	0.329

Table 2. ANCOVA post-hoc pairwise comparisons for Pre-to Post-test mean accuracies. SE stands for Standard Error. $^{**}p < 0.01$, $^{*}p < 0.05$. Bonferroni Corrections were conducted for multiple pairwise comparisons.

Condition	Category	Surv.	Pre vs. Post	N	μ	σ	t	p	What's Tested?
Comparison	Omelette	Y	Repeated	32	28.375	71.922	2.232	0.033*	Learning
	Salad_G	Y	Repeated		17.125	79.246	1.222	0.231	Learning
	Yogurt	Y	Repeated		16.500	47.268	1.975	0.057	Learning
	Pancake	N	Repeated		21.156	56.781	2.108	0.043*	Learning+Transfer
	Salad_O	N	New		29.161	62.473	2.599	0.014*	Transfer
	Meat_T	N	New		-2.656	34.950	-0.430	0.670	Transfer
Decomposition	Omelette	Y	Repeated	31	14.129	49.574	1.587	0.123	Learning
	Salad_G	Y	Repeated		15.065	91.246	0.919	0.365	Learning
	Yogurt	Y	Repeated		5.500	14.339	2.136	0.041*	Learning
	Pancake	N	Repeated		12.387	50.156	1.375	0.179	Learning+Transfer
	Salad_S	N	New		16.871	39.054	2.405	0.023*	Transfer
	Meat_C	N	New		19.210	76.278	1.402	0.171	Transfer

Table 3. The pre- and post-test means paired samples t-test results. 'Category' denotes the type of meal in the photos (Salad_G: Greek Salad, Salad_O: Original Salad, Salad_S: Sabzi Salad, Meat_T: Tilapia, Meat_C: Cod); 'Surv.' denotes if the meal photo was used in the the survey (Y) or not (N); 'Pre vs. Post' denotes whether the meal photo at pre-test is repeated at post-test; and 'What's tested?' denotes what type of learning occurred. $^{*}p < .05$. Bonferroni Corrections were conducted for multiple pairwise comparisons.

estimation. We collected those differences and took the mean of them for aggregate analysis (see Table 1, first four columns). Overall, there were no statistically significant differences in the means between the Comparison and Decomposition groups' estimates (t = 0.520, p = 0.603), but Comparison and Decomposition did better than the Control group. Comparison showed approaching significance compared to Control (t = -1.845, p = 0.066), while Decomposition showed a statistically significant difference compared to Control (t = -2.362, p = 0.019).

4.3.3 Timing of Survey Completion. The Comparison group took a total of 51389 seconds to complete the survey as a group and when it was divided by the number of participants (n=31), the average was 1657.71 seconds (27.63 minutes) per participant. The Decomposition group took 59754 seconds to complete the survey and an average of 1927.55 seconds (32.13 minutes) per participant. Therefore, on average, the Comparison group appeared to finish the survey activity faster than the Decomposition group. However, when an independent samples t test was conducted between the two group's time completion for all participants, there was no statistically significant difference (t=-0.894, p=0.375).

4.3.4 Evidence for Carb Estimation Learning. We looked at the difference in carb estimation accuracy from pre- to post-tests for all conditions. For this metric, we subtracted participants' accuracy for the post-test, calculated using the same method as described above, from their accuracy for pre-tests using the following formula: $(|Post - GT|) - (|Pre - GT|)$; this was calculated for all six meal categories in pre- and post-tests for all conditions, respectively. If a participant estimated the carb content more accurately at post-test, this metric would return a greater negative value. If their accuracy was lower at pre-test, it would return a greater positive value (see Table 1, last three columns). We used ANCOVA to examine if the change from pre- to post-test was different among the three conditions (see Table 2). Bonferroni Corrections were conducted for multiple pairwise comparisons. Overall, there was a statistically significant difference in the means of estimates among the three groups (F(2, 554)=5.024, p=.007, η_p^2=.018), adjusting for education level.

In order to determine if learning and transfer of knowledge occurred, some questions in the pre-test were repeated (i.e., people were given the same meal photographs) in the post-test, while others were not repeated (i.e., only new meal photographs used in the post-test). Similarly, some meal photographs were only used in the pre- and post-tests but not during the experimental activity phase (denoted through Y or N under "Surv" in Table 3). We stated that learning occurred if there was an increase in nutritional estimation accuracy for any repeated meal photos. Transfer of knowledge occurred if there was an increase in nutritional estimation accuracy for similar but different and new meal photos. Bonferroni Corrections were conducted for multiple pairwise comparisons.

Participants in the Comparison Strategy, showed learning from pre- to post-test for the "Omelette" and "Pancake," categories, while transfer of knowledge occurred in the "Pancake" and "Original Salad" categories. Participants in the Decomposition Strategy, showed learning from pre- to post-test for the "Yogurt" category, while transfer of knowledge occurred in the "Sabzi Salad" category as shown in Table 3.

4.3.5 User Feedback About the Experience. There were 3 themes identified from the analysis of the comments from the experimental groups: 1) Inspiration for healthy eating, 2) Getting through the learning curve, and 3) Questions about accuracy.

[1] *Inspiration for Healthy Eating.* Participants in both the Comparison and Decomposition groups expressed how the experience inspired them to try new foods and eat healthier.

Comparison group participants made statements such as:

- *"It makes you want to try different foods."*
- *"It could be a great way to get recipe ideas if you are trying to stay in a certain range."*
- *"It helps to have a variety of recipes that I could never have imagined."*

Similarly, the Decomposition group stated:

- *"It does show that there can be some variety in low-carb eating, which many people think low-carb eating lacks."*
- *"Seeing meals prepared and posted by others is helpful in giving ideas for healthier eating."*

[2] *Getting Through the Learning Curve.* Both experimental groups discussed the learning curve that these nutritional estimation strategies require. Participants from the Comparison groups made reflections such as:

- *"The more I did the better I got...with enough practice that I could get pretty good doing it this way."*
- *"With time it can become a really solid method for people to understand carbs in their food. By using this way to learn, I slowly came to learn types of foods that had more carbs than others. By the end of the experiment, I didn't make the best judgements, but if I had more time with the system I'm sure I could be quite precise."*

Comments from the Decomposition groups echoed these sentiments:

- *"It started to become easier to remember which foods were about how many carbs."*
- *"...it is good, a little hard, but I got the hang of it at the end."*
- *"Only relying on a picture to assess nutrition in meals would be hard when you first started, but with more time could become easier for people to learn and use more often. I also think that as*

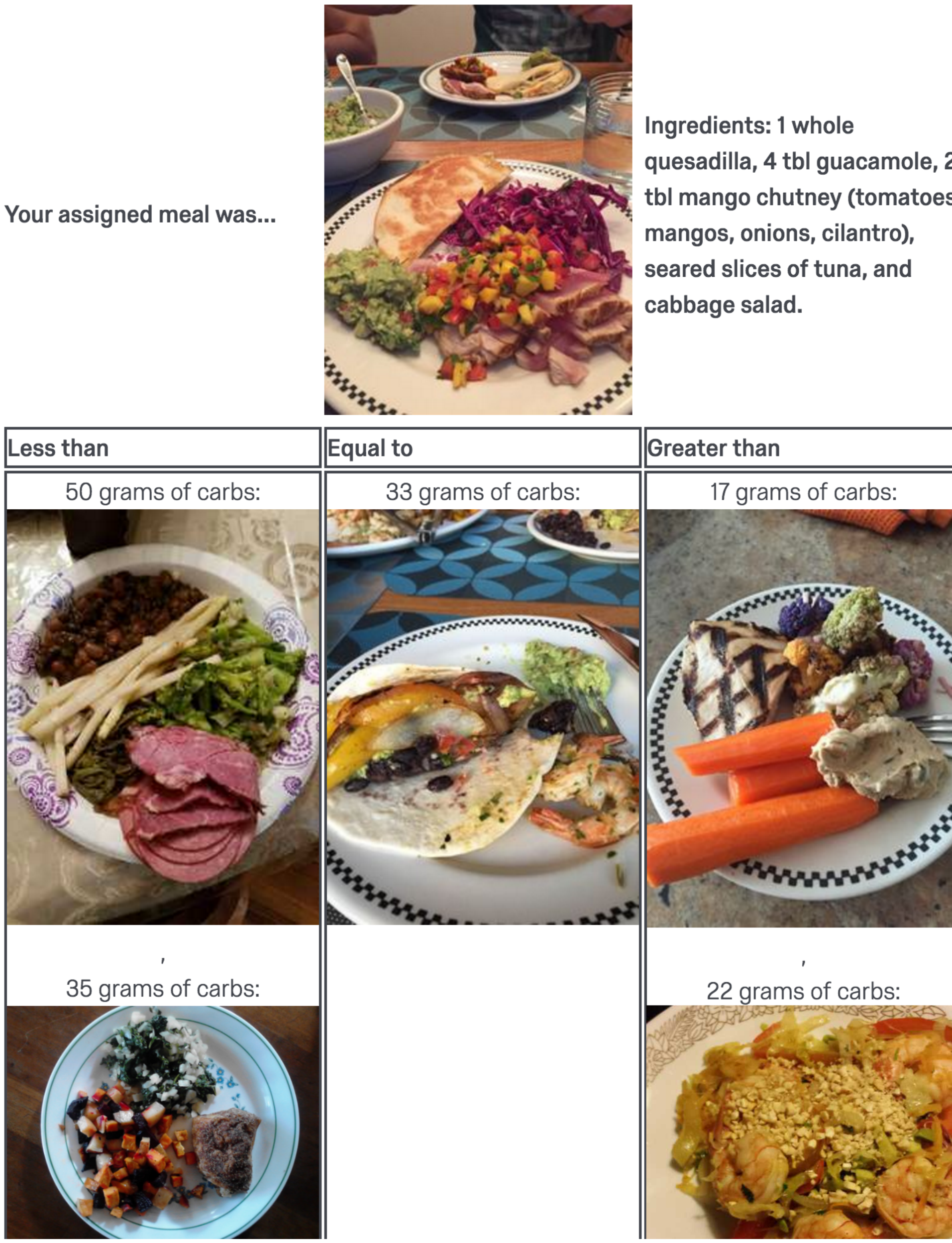

Fig. 4. Sample picture of a user's selection summary at the end of a round in the Comparison Strategy group.

> *people begin to learn the different foods and the amount of carbs in them they would soon be assessing more of their meals and make healthier choices."*

Overall, the majority of participants in the experimental groups (68% Decomposition, 70% Comparison) rated their respective strategies as useful and effective. This question was not asked to the Control group as there was not a particular strategy that was involved in their tasks. However, for the Control group, the majority of comments stated that the the assigned task was "time consuming" and "incredibly difficult".

[3] *Questions about Accuracy.* Both experimental groups were concerned about the accuracy of their estimations. The Decomposition group noted more concerns about the accuracy specifically:

- *"It would be better with more accurate units of measurement and conversion between different units."*
- *"Anyone truly trying to keep serious counts wouldn't bother with this."*

For the Comparison group, negative comments remarked that the approach was a rough estimate, but could nonetheless provide benefits, e.g.:

- *"A good way to get a rough estimate of what a meal is like based on another meal."*
- *"An interesting way of evaluating nutritional value and content. I don't feel you get the exact scope of what the meal contains, but it's still a good metric to look at and estimate in this way."*
- *"An educational way of looking at the nutritional values."*

4.3.6 Crowdsourcing Results. Between the Decomposition and Comparison groups, five non-evaluated meal images were nutritionally assessed 62 times per meal, aggregating to 310 unique assessments. With the Control group, that number increased to 465 assessments. For each round, the non-evaluated meal was compared to five different but similar meal images (such as Fig. 3 left image), resulting in 25 meal images that were labeled as less, more, or the same amount of carbs to the comparison meal (such as Fig. 4 bottom five images). All of these crowdsourced nutritional assessments were completed in 4-5 minutes per participant per meal image.

5 DISCUSSION

5.1 Study 1: Can People Learn Nutrition Estimation from the Crowd?

Two key results are striking from Study 1. First, the CB did not appear to support learning of nutrition estimation skills. Second, users appeared to agree with the CB more often than they reported being influenced by the CB to change their answer.

5.1.1 Untrained CB Alone Does Not Support Learning. Several factors may have contributed to the finding that access to an untrained CB did not support learning. The most likely factor is related to the accuracy of the CB. The CB results were generated from lay users and did not always result in the correct meal as the most highly voted meal. If users trusted the community's opinion, but then found out that their original meal choice was correct, this could result in confusion and trust issues with the crowd. These results align with a 2017 study by Burgermaster et al. [9] that examined how exposure to visually depicted proportions of the responses from other participants (i.e. peer comparison) impacted macronutrient estimation learning via meal photos. This study found that the expert feedback condition (i.e. correctness information with explanation why) led to significantly higher post-test accuracy than the other conditions of peer comparison, expert explanation (i.e. only correctness information), and control. However, they also found that the difference between expert and peer comparison was not significant. Thus, our study built on this prior work by including the crowds' reasoning along with their selection percentage data 1 (a) and (b), but found that this addition did not significantly impact learning. Given this finding, one possible direction for future work would be to test whether the inaccurate crowd issue still holds with a larger crowd size. It is possible that more responses might improve the crowd's overall wisdom as the weight of any outlier incorrect responses diminish. However, it is unclear whether adding more lay users would mitigate this issue.

Given our results, a more promising potential solution is for future apps with CB-like components to increase the accuracy of crowd wisdom by recruiting from a specific population (e.g. registered dietitians, or even simply members of the public with carb counting experience) to generate the initial pool of CB responses. A 2015 study by Abbar et al. showed that Twitter and US Census data were sensitive enough to pick out personal dietary variations based on users' interests, demographics, and social networks. Results showed differences in gender, education level, and users' self-disclosed habits related to their dietary habits. Future work could use similar approaches to identify which characteristics predict the completion of accurate macronutrient assessment of meal photos, such that people fitting these characteristics could be recruited to work on the CB responses [1]. If done correctly, this approach could potentially skew the likelihood that the correct responses would be the most highly voted. Another approach to build up the accuracy of the crowd that could be explored is the application of techniques that promote the accuracy of crowdsourced responses such as peer-agreement and gold checks [35] or augmentation systems that rely on complex workflows to improve crowd accuracy [28] such as *Cicero* that was developed by Chen et al. in 2019 [17].

It is also possible that the cognitive load for users in Study 1 was to high to facilitate learning. Our users had to review information about meal photos, but also interpret the crowd's evaluations as well. Prior literature has pointed to the difficulty in estimating the nutrition content of meals in general. It is thus possible that our solution, while created to be as easy and lightweight as possible, did not reduce enough of the mental burden to facilitate arrival at the best answer choice. Alternatively, having every user review the CB for every round could have resulted in people engaging less with thinking through their responses during each round and waiting for the CB results to decide for them. While our results regarding the CB influence did not point to this occurring for many users (discussed further below in 5.1.2), examining the potential that some users took less ownership of their own learning, resulting in lower recall scores, due to the knowledge that the community would offer an answer is another possible direction for future work.

5.1.2 CB Agreement vs. CB Influence. The second interesting result from Study 1 is related to the degree of overall influence of the CB. Even though a high percentage of users agreed with the CB, the report on how that influenced their final answer choice was much lower. Hence, just agreeing with the community's opinion did not necessarily lead to the specific action of changing their answers to match the CB. About half of the users never changed their answers for all five rounds, which suggests either they 1) were not influenced by the crowd's opinion or 2) found their answers were already in agreement with the CB's responses.

It is unclear in the former scenario as to why users did not appear to be influenced by the crowd's opinion enough to change their answers. It is possible that while users found themselves not disagreeing with the crowd's selection and reasoning, they simply had more confidence in their own responses. Prior research has focused on clarifying the role of social influence on eating behaviors by first establishing that unhealthy behaviors appear to be contagious [81] with subsequent work examining if observed similarities are attributable to social influence or if it is actually people's tendency to form ties with similar people who have similar habits. A 2021 study by Gligorić et al. found evidence of a clear social influence for people who acquired a healthy-eating partner resulting in significantly healthier eating. This study supports the premise that social influences are present in food decisions that could ultimately be applied to future work seeking to better leverage the crowd's influence in nutrition estimation [34]. Further, a study by Pater et al. from 2016 explored eating disorders in social media spaces and raised salient points regarding the impact that technology combined with the influence of the crowd can have on behavior [67]. This has implications for our work since we were similarly examining the role of crowd influence within an anonymous technology-based platform, which might have allowed our users to detach from the activity and discount the wisdom of the crowd. A future study might seek to better understand the mental process that users take when considering peer responses and explanations, including scenarios in which they consider changing their answer but ultimately decide against it.

In the second study from the 2017 Burgermaster et al. [9] paper, the authors tested expert explanations against peer explanations in nutrition estimation of meal photos, and found both conditions led to similar gains in learning compared to controls. The authors noted, however, that there were notable differences in quality and helpfulness of peer-generated explanations when compared to the relatively consistent explanations offered by experts. This speaks to further exploring what elements are critical to high quality explanations for both actual learning and related to crowd influence. A final interesting factor worth examining in more depth is the role that information about the crowd has on the level of influence for the users' answer choice in this context. For example, telling users key information about the crowd such as relevant expertise or credentials would likely change the level of influence that crowd has, but how would users respond knowing the crowd was comprised of people selected for a different reason such as general education attainment or predicted ability to be accurate at nutrition estimation? Further, how would the influence of the crowd wane if users were told that it was comprised of experts, but found it to be repeatedly incorrect?

5.2 Study 2: Assessing Strategies for Learning Carbohydrate Estimation through Digital Intervention

In Study 2, we compared the impact of two different strategies (Decomposition and Comparison) for teaching individuals the skills they need to arrive at efficient and accurate nutrition estimates of their meals.

5.2.1 Decomposition Strategy Findings. The Decomposition approach helped individuals map out the mental process of nutrition estimation into concrete steps, and provided them with assistance on these steps (i.e. providing them with a macronutrient reference table). On the surface, this approach appears to have a number of benefits. For example, it does not rely on availability of any additional data beyond a single user's record and, as such, is more easily scalable. Moreover, it is consistent with empirically validated existing methods for improving nutritional assessment of meals and with traditional teaching practices of diabetes educators.

However, this approach also presents challenges scaling to complex meals with multiple ingredients. While the participants in our study reported having a generally positive experience with the Decomposition assistance tool, this could change if the meals included 10 or more ingredients, as is common for (e.g. mixed salads). It is relevant to mention that the dietitians who generated gold standard nutritional assessments used in this study sometimes took more than 10 minutes to arrive at a thorough assessment. Users may not be willing to spend 10 minutes in front of a meal working through its nutritional assessment.

5.2.2 Comparison Strategy Findings. The Comparison Strategy was designed to be lightweight relative to the Decomposition strategy. Participants were asked to simply compare pairs of meals side-by-side, without having to rely on an assistance tool or additional complexities. Moreover, it appears to have an additional benefit of exposing an individual to a variety of different comparison meals captured by others, thus broadening the range of their potential nutritional choices.

Interestingly, users showed learning across three types of food categories when using the Comparison strategy (as opposed to two types when using Decomposition). Additionally, users stated that while the Comparison approach did not result in perfectly accurate assessments, it was still interesting, useful to consider, and/or educational. The Comparison strategy thus appears to at least be on par with the traditional Decomposition strategy, and may even be more beneficial for users in some respects.

We note, too, that the Comparison strategy has potential limitations. For example, it relies on the availability of labeled sets of meal images with their nutritional estimation. It also may be sensitive to sub-optimal choices of similar meals during its first step. In Study 2, we circumvented this challenge by always presenting participants with the best-fitting comparison meal for the target meal. This feature is currently not available to a sufficiently scalable degree in any real-world app. The degree to which a sub-optimal first choice may negatively impact individuals' ability to compare meals on their nutrition is still unclear. Future work might examine this factor as well as seek to minimize this burden on users.

5.2.3 A Closer Look: Decomposition vs. Comparison. Future studies will need to assess the finer details of the potential effect of education level on the capacity for nutritional estimation skill learning, as prior studies have seen education as predictors for decrease in obesity percentage and sharing consumption of low caloric food on social media but not for nutrition skills as well [1]. Broadly, the results from Study 2 showed that the experimental strategies appeared to support individuals in terms of (1) their accuracy in estimation and (2) their ability to estimate future meals as compared to a control. This suggests that both methods may help promote nutritional learning for the average user. More specifically, the Decomposition approach had a statistically significant impact on the accuracy of nutritional estimation during the experimental activity, while the Comparison approach did not. This is not surprising, as the former group had the macronutrient reference table to guide their carbohydrate estimation of each meal. If the user was willing to put time and calculation into the task, accurate estimation was always possible for the former group, while the Comparison group always had to guess and make a judgment call. However, intriguing results were observed when we compared the two groups on their learning and knowledge transfer. Though both strategies showed statistical significance for learning and knowledge transfer, the Comparison group performed considerably better than the Decomposition group.

A salient question that remains is how to ensure the dataset needed for a Comparison-based approach is sufficiently (1) accurate and (2) comprehensive. While some meal records on social media platforms contain nutritional information, most do not. Leveraging crowdsourcing could be a potential solution, given the extent of data we were able to collect within a small timeframe. However, if crowdsourcing is used in the future to collect this kind of data, the accuracy of the labels must be ensured. One direction for future work could be to combine crowdsourced intelligence from trained users (human-in-the-loop) [59] with machine learning approaches [65]. Once users are trained well enough with the Comparison approach they would be well-equipped to continue providing labels. A 2016 study by Chancellor et al. illustrates how social media posts have been used to develop predictive algorithms for mental illness severity in people posting pro-eating disorder content. Future work could similarly leverage tagged photos from social media to develop a machine learning model that generates a diverse nutrition database of meal photos [11]. Furthermore, efforts could be made to combine prior HCI research efforts that have meal photo repositories towards generating a large, diverse database [27, 31, 52, 54]. Given that gamification has been suggested to improve crowdsourced data quantity and quality in the context of mobile apps [83], a game environment (e.g., as in *Carb Counting with Lenny*) might be a suitable approach. A higher quality food database could then be achieved by creating a symbiotic relationship between intelligent algorithms and educated users, leading to healthier eating overall. Finally, a different approach to extracting the nutritional information used by Munmun De Choudhury et al. [30] reported a matching method that had been used in other research with 89% accuracy [73], where Instagram posts of tagged meal photos were matched to the USDA food descriptors. The authors found they were able to extract nutrition information for 93.5% of the Instagram posts in their set [24]. However, the focus of this work was not on the accuracy of the nutritional profile, but instead was on using this profile in their analysis of the types of foods eaten in food deserts versus non-food deserts. Future work should evaluate the application of this matching technique for accuracy of the macronutrient evaluations themselves.

The overall positive qualitative comments were promising for the two experimental strategies. Future work might include assessing the same strategies within a volunteer space to determine if compensation incentives affected the results. Some of the negative comments across all conditions included that the task was "unnecessarily obtuse and confusing," "difficult," and that it has a "learning curve." A future study could address the learning curve that users identified by seeking to make nutritional estimation even easier and more accessible. For example, a future interface might include the implementation of a Tinder-like swiping mechanism to give users more exposure and practice to these estimation approaches more rapidly. The addition of this mechanism might make it easier for lay users to provide data. It may also help in aggregating a larger set of data of meal photos with labels that indicate their comparable macronutrient amounts. Future diet tracking apps could be designed to use a similar swiping mechanism as part of either the Comparison or Decomposition approach. Further reducing the complexity and time commitment to macronutrient estimation would potentially address the known relevant barriers identified from the literature on capturing meal photos/food journaling [21, 22, 29]. There is already an example of breaking down the steps for nutritional evaluation [13] which is consistent with the Decomposition strategy and could facilitate future apps in this way. However, applying the Comparison approach in this way with user-generated meal photos has not been tried before and would need to be evaluated in the context of dietary tracking and self-monitoring apps.

6 LESSONS LEARNED AND FUTURE DIRECTIONS

Our results suggest the following key lessons learned along with ideas for future work:

(1) **The crowd doesn't necessarily teach macronutrient assessment**: Users did not appear to reliably learn macronutrient assessment through viewing community decisions and justifications. In fact, the crowd may harm learning in certain contexts. Nutrition education apps with a crowd component should take care to evaluate accuracy and trust in their situational contexts. Future work should examine if there is a subset of the population that would find this approach to be the preferred way to learn nutrition estimation and find ways to improve the accuracy of the crowd.

(2) **Carb estimation skills need an educational component and can be improved via the Decomposition or Comparison approaches**: Users can build carbohydrate estimation skills which they find useful by learning either the Decomposition or Comparison strategy. Future work needs to further examine these educational strategies in larger groups and consider how to further scaffold the process during early use to address the reported learning curve phenomenon.

(3) **Crowdsourcing meal photos and labels from app-trained experts may be a powerful approach to support future macronutrient assessment education and inspire healthier eating**: Through Study 2, 93 users created 465 unique assessments of meal photos with an average time of less than five minutes per evaluation. Given that many users found the experience useful, this (or a related) approach may be a viable way to crowdsource new photos - and perhaps even corresponding labels - to support future work in nutrition and health if implemented at scale. Additionally, users in this study indicated they were inspired to try new foods and eat healthier by viewing someone else's labeled meal photos as part of learning carbohydrate estimation skills.

On the whole, the results in the studies presented here suggest promise for the use of a lightweight way to teach macronutrient estimation skills via digital intervention. Future apps which are able to successfully leverage the findings from lesson 3 may be able to encourage users to more regularly practice their nutrient assessment skills, discover new healthy options that work best for them, and enjoy the process overall. Given ideas 1 and 3, a future application which supports dietary health might choose to include meal photos related to health goals from other users but not necessarily highlight their estimation and reasoning results for others to view. Research going forward should examine the possibility of leveraging crowdsourcing with app-assisted nutritional literacy education and examine how more nutritional literacy skills might be best learned.

ACKNOWLEDGMENTS

This work was funded in part by the NSF Smart and Connected Health (Grant No. 1551708, PI Mamykina) as well as NSF Human-Centered Computing (Grant no. 2104515, PI Hwang).

REFERENCES

[1] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*. 3197–3206.

[2] Sharmini Abdullah, Mohamed Elshaikh, Muhammad Bazli Mahmood, Norshahrizan Nordin, Syed Zulkarnain Syed Idrus, Rozilawati Mahadi, and Umie Naiemah Saraih. 2018. Food Code Breaker (FCB)–Developing the Multi-Lingual Food Code Translation Android Application Using Multi–Options Code Reader System. In *Journal of Physics: Conference Series*, Vol. 1019. IOP Publishing, 012013.

[3] Ala'a Alkerwi, Georgina E Crichton, and James R Hébert. 2015. Consumption of ready-made meals and increased risk of obesity: findings from the Observation of Cardiovascular Risk Factors in Luxembourg (ORISCAV-LUX) study. *British Journal of Nutrition* 113, 2 (2015), 270–277.

[4] Sara Altamirano, Gudrun Thorsteinsdottir, and Verónica Burriel. 2019. Mobile application for celiac disease patients' wellness and support. In *International Conference on Wireless Mobile Communication and Healthcare*. Springer, 18–35.

[5] Marios Anthimopoulos, Joachim Dehais, Sergey Shevchik, Botwey H Ransford, David Duke, Peter Diem, and Stavroula Mougiakakou. 2015. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. *Journal of diabetes science and technology* 9, 3 (2015), 507–515.

[6] American Diabetes Association. 2021. Carb Counting and Diabetes. https://diabetes.org/healthy-living/recipes-nutrition/understanding-carbs/carb-counting-and-diabetes.

[7] Marcela CC Bomfim, Sharon I Kirkpatrick, Lennart E Nacke, and James R Wallace. 2020. Food Literacy while Shopping: Motivating Informed Food Purchasing Behaviour with a Situated Gameful App. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[8] Marcela CC Bomfim and James R Wallace. 2018. Pirate Bri's Grocery Adventure: Teaching Food Literacy through Shopping. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[9] Marissa Burgermaster, Krzysztof Z Gajos, Patricia Davidson, and Lena Mamykina. 2017. The Role of Explanations in Casual Observational Learning about Nutrition. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4097–4145.

[10] Sarah Campos, Juliana Doxey, and David Hammond. 2011. Nutrition labels on pre-packaged foods: a systematic review. *Public health nutrition* 14, 8 (2011), 1496–1506.

[11] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1171–1184.

[12] Pierre Chandon and Brian Wansink. 2007. The biasing health halos of fast-food restaurant health claims: lower calorie estimates and higher side-dish consumption intentions. *Journal of Consumer Research* 34, 3 (2007), 301–314.

[13] Pierre Chandon and Brian Wansink. 2007. Is obesity caused by calorie underestimation? A psychophysical model of meal size estimation. *Journal of Marketing Research* 44, 1 (2007), 84–99.

[14] Beenish M Chaudhry, Christopher Schaefbauer, Ben Jelen, Katie A Siek, and Kay Connelly. 2016. Evaluation of a Food Portion Size Estimation Interface for a Varying Literacy Population. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5645–5657.

[15] Beenish Moalla Chaudry, Kay Connelly, Katie A Siek, and Janet L Welch. 2013. Formative evaluation of a mobile liquid portion size estimation interface for people with varying literacy skills. *Journal of ambient intelligence and humanized computing* 4, 6 (2013), 779–789.

[16] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh fast-food image dataset. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 289–292.

[17] Quanze Chen, Jonathan Bragg, Lydia B Chilton, and Dan S Weld. 2019. Cicero: Multi-turn, contextual argumentation for accurate crowdsourcing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[18] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1674–1687.

[19] GBD 2015 Obesity Collaborators. 2017. Health effects of overweight and obesity in 195 countries over 25 years. *New England Journal of Medicine* 377, 1 (2017), 13–27.

[20] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3207–3216.

[21] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3207–3216. https://doi.org/10.1145/2702123.2702154

[22] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. 2015. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1159–1162.

[23] Helena Dória Ribeiro de Andrade Previato. 2016. Carbohydrate Counting in Diabetes. *Nutrition and Food Technology: Open Access* (2016). http://dx.doi.org/10.16966/2470-6086.124

[24] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. 2016. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 1157–1170.

[25] Joachim Dehais, Marios Anthimopoulos, and Stavroula Mougiakakou. 2016. GoCARB: A Smartphone Application for Automatic Assessment of Carbohydrate Intake. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. ACM, 91–91.

[26] Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, and Stavroula Mougiakakou. 2017. Two-view 3D Reconstruction for Food Volume Estimation. *IEEE transactions on multimedia* 19, 5 (2017), 1090–1099.

[27] Pooja M Desai, Elliot G Mitchell, Maria L Hwang, Matthew E Levine, David J Albers, and Lena Mamykina. 2019. Personal health oracle: Explorations of personalized predictions in diabetes self-management. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[28] Ryan Drapeau, Lydia Chilton, Jonathan Bragg, and Daniel Weld. 2016. Microtalk: Using argumentation to improve crowdsourcing accuracy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 4.

[29] Daniel A Epstein, Felicia Cordeiro, James Fogarty, Gary Hsieh, and Sean A Munson. 2016. Crumbs: lightweight daily food challenges to promote engagement and mindfulness. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5632–5644.

[30] Daniel Fried, Mihai Surdeanu, Stephen Kobourov, Melanie Hingle, and Dane Bell. 2014. Analyzing the language of food on social media. In *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 778–783.

[31] Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells, Jonathan N. Tobin, Esteban G. Tabak, et al. 2021. From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition Goal Recommendations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[32] Adriana Servilha Gandolfo, Débora Vasconcelos Bastos, Bruna Abreu Jabur Makluf, Lenycia de Cassya Lopes Neri, Roberta Diaz Savoldelli, Thais Della Manna, Durval Damiani, Alexandre Archanjo Ferraro, et al. 2014. Efficacy of photographic educational materials for carbohydrate counting training of adolescents with diabetes mellitus. *Nutricion hospitalaria* (2014).

[33] Sandra J Gillespie, Karmeen D Kulkarni, and Anne E Daly. 1998. Using carbohydrate counting in diabetes clinical practice. *Journal of the American Dietetic association* 98, 8 (1998), 897–905.

[34] Kristina Gligorić, Ryen W White, Emre Kiciman, Eric Horvitz, Arnaud Chiolero, and Robert West. 2021. Formation of social ties influences food choice: A campus-wide longitudinal study. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[35] Tanya Goyal, Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, and Matthew Lease. 2018. Your behavior signals your reliability: Modeling crowd behavioral traces to ensure quality relevance annotations. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

[36] Andrea Grimes and Rebecca E Grinter. 2007. Designing persuasion: Health technology for low-income African American communities. In *International Conference on Persuasive Technology*. Springer, 24–35.

[37] Andrea Grimes, Vasudhara Kantroo, and Rebecca E Grinter. 2010. Let's play!: mobile health games for adults. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 241–250.

[38] Rohana N Haththotuwa, Chandrika N Wijeyaratne, and Upul Senarath. 2020. Worldwide epidemic of obesity. In *Obesity and obstetrics*. Elsevier, 3–8.

[39] Jackie Haven, Adam Burns, Patricia Britten, and Carole Davis. 2006. Developing the consumer interface for the MyPyramid food guidance system. *Journal of nutrition education and behavior* 38, 6 (2006), S124–S135.

[40] Simo Johannes Hosio, Niels van Berkel, Jonas Oppenlaender, and Jorge Goncalves. 2020. Crowdsourcing personalized weight loss diets. *Computer* 53, 1 (2020), 63–71.

[41] Mary Margaret Huizinga, Adam J Carlisle, Kerri L Cavanaugh, Dianne L Davis, Rebecca P Gregory, David G Schlundt, and Russell L Rothman. 2009. Literacy, numeracy, and portion-size estimation skills. *American journal of preventive medicine* 36, 4 (2009), 324–328.

[42] Maria L Hwang and Lena Mamykina. 2017. Monster appetite: Effects of subversive framing on nutritional choices in a digital game environment. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4082–4096.

[43] Mina C Johnson-Glenberg and Eric B Hekler. 2013. " Alien Health Game": An Embodied Exergame to Instruct in Nutrition and MyPlate. *Games for Health: Research, Development, and Clinical Applications* 2, 6 (2013), 354–361.

[44] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. 2016. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association* 23, 3 (2016), 440–448.

[45] Whitley R Kemble. 2014. *A proposed model for designing children's health-focused serious games*. Ph.D. Dissertation. Iowa State University.

[46] David A Kindig, Allison M Panzer, Lynn Nielsen-Bohlman, et al. 2004. *Health literacy: a prescription to end confusion*. National Academies Press.

[47] Fanyu Kong and Jindong Tan. 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing* 8, 1 (2012), 147–163.

[48] Pallavi Kuhad, Abdulsalam Yassine, and Shervin Shimohammadi. 2015. Using distance estimation and deep learning to simplify calibration in food calorie measurement. In *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2015 IEEE International Conference on*. IEEE, 1–6.

[49] Karmeen D Kulkarni. 2005. Carbohydrate counting: a practical meal-planning option for people with diabetes. *Clinical Diabetes* 23, 3 (2005), 120–122.

[50] Charalampos Kyfonidis and Marilyn Lennon. 2019. Making diabetes education interactive: tangible educational toys for children with type-1 diabetes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[51] M Leveridge, A Loomes, J Urwin, and S Zahra. 2017. Cook and Eat Exeter: Carbohydrate counting in real life. *Diabetes Care for Children & Young People* 7 (2017).

[52] Brian Y Lim, Xinni Chng, and Shengdong Zhao. 2017. Trade-off between automation and accuracy in mobile photo recognition food logging. In *Proceedings of the Fifth International Symposium of Chinese CHI*. 53–59.

[53] Medtronic, Inc. [n.d.]. Carb Counting with Lenny. https://apps.apple.com/us/app/carb-counting-with-lenny/id516080517

[54] Michele Merler, Hui Wu, Rosario Uceda-Sosa, Quoc-Bao Nguyen, and John R Smith. 2016. Snap, Eat, RepEat: a food recognition engine for dietary logging. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. ACM, 31–40.

[55] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. 2015. Im2Calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*. 1233–1241.

[56] Tatsuya Miyazaki, Gamhewage C de Silva, and Kiyoharu Aizawa. 2011. Image-based calorie content estimation for dietary assessment. In *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 363–368.

[57] Barry Mulrooney, Mairéad McDermott, and Nick Earley. 2006. NutraStick: portable diet assistant. In *CHI'06 extended abstracts on Human factors in computing systems*. 1855–1860.

[58] Michael Nelson, M Atkinson, and S Darbyshire. 1996. Food photography II: use of food photographs for estimating portion size and the nutrient content of meals. *British Journal of Nutrition* 76, 01 (1996), 31–49.

[59] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos. 2011. Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 1–12.

[60] Francisco Nunes, João Almeida, Chia-Fang Chung, and Nervo Verdezoto. 2021. Avoiding Reactions Outside the Home: Challenges, Strategies, and Opportunities to Enhance Dining Out Experiences of People with Food Hypersensitivities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[61] U.S. Department of Agriculture. 2021. What is MyPlate? https://www.myplate.gov/eat-healthy/what-is-myplate.

[62] U.S. Department of Health, Centers for Disease Control Human Services, and & National Center for Health Statistics Prevention. 2021. Obesity, Race/Ethnicity, and COVID-19, 2011–2014. Retrieved August 11, 2021 from https://www.cdc.gov/obesity/data/obesity-and-covid-19.html.

[63] Samantha Olinsky, Pooja M Desai, Selen Turkay, Elizabeth M Heitkemper, Elliot G Mitchell, Lena Mamykina, and Maria L Hwang. 2021. Meals for Monsters: a Mobile Application for the Feasibility of Gaming and Social Mechanisms. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[64] Rita Orji, Julita Vassileva, and Regan L Mandryk. 2013. LunchTime: a slow-casual game for long-term dietary behavior change. *Personal and Ubiquitous Computing* 17, 6 (2013), 1211–1221.

[65] Tariq Osman Andersen, Francisco Nunes, Lauren Wilcox, Elizabeth Kaziunas, Stina Matthiesen, and Farah Magrabi. 2021. Realizing AI in Healthcare: Challenges Appearing in the Wild. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.

[66] Joongsin Park, Bon-chang Koo, Jundong Cho, and Byung-Chull Bae. 2015. Snackbreaker: A game promoting healthy choice of snack foods. In *Proceedings of the 2015 annual symposium on computer-human interaction in play*. 673–678.

[67] Jessica A Pater, Oliver L Haimson, Nazanin Andalibi, and Elizabeth D Mynatt. 2016. "Hunger Hurts but Starving Works" Characterizing the Presentation of Eating Disorders Online. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1185–1200.

[68] Parisa Pouladzadeh, Pallavi Kuhad, Sri Vijay Bharat Peddi, Abdulsalam Yassine, and Shervin Shirmohammadi. 2016. Food calorie measurement using deep learning neural network. In *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International*. IEEE, 1–6.

[69] Daniel Rhyner, Hannah Loher, Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, Ransford Henry Botwey, David Duke, Christoph Stettler, Peter Diem, and Stavroula Mougiakakou. 2016. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. *Journal of medical Internet research* 18, 5 (2016).

[70] Russell L Rothman, Ryan Housam, Hilary Weiss, Dianne Davis, Rebecca Gregory, Tebeb Gebretsadik, Ayumi Shintani, and Tom A Elasy. 2006. Patient understanding of food labels: the role of literacy and numeracy. *American journal of preventive medicine* 31, 5 (2006), 391–398.

[71] Hanna Schäfer, Mehdi Elahi, David Elsweiler, Georg Groh, Morgan Harvey, Bernd Ludwig, Francesco Ricci, and Alan Said. 2017. User nutrition modelling and recommendation: Balancing simplicity and complexity. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*. 93–96.

[72] Dean Schillinger, Kevin Grumbach, John Piette, Frances Wang, Dennis Osmond, Carolyn Daher, Jorge Palacios, Gabriela Diaz Sullivan, and Andrew B Bindman. 2002. Association of health literacy with diabetes outcomes. *Jama* 288, 4 (2002), 475–482.

[73] Sanket S Sharma and Munmun De Choudhury. 2015. Measuring and characterizing nutritional information of food and ingestion content in instagram. In *Proceedings of the 24th International Conference on World Wide Web*. 115–116.

[74] Nancy F Sheard, Nathaniel G Clark, Janette C Brand-Miller, Marion J Franz, F Xavier Pi-Sunyer, Elizabeth Mayer-Davis, Karmeen Kulkarni, and Patti Geil. 2004. Dietary carbohydrate (amount and type) in the prevention and management of diabetes. *Diabetes care* 27, 9 (2004), 2266–2271.

[75] Katie A Siek, Kay H Connelly, Beenish Chaudry, Desiree Lambert, and Janet L Welch. 2009. Evaluation of two mobile nutrition tracking applications for chronically ill populations with low literacy skills. *Mobile Health Solutions for Biomedical Applications* (2009), 1–23.

[76] Katie A Siek, Kay H Connelly, and Yvonne Rogers. 2006. Pride and prejudice: learning how chronically ill people think about food. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 947–950.

[77] Katie A Siek, Kay H Connelly, Yvonne Rogers, Paul Rohwer, Desiree Lambert, and Janet L Welch. 2006. When do we eat? An evaluation of food items input into an electronic food monitoring application. In *Pervasive Health Conference and Workshops, 2006*. IEEE, 1–10.

[78] Kami J Silk, John Sherry, Brian Winn, Nicole Keesecker, Mildred A Horodynski, and Aylin Sayir. 2008. Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. *Journal of nutrition education and behavior* 40, 1 (2008), 3–10.

[79] D Lopes Souto and E Lopes Rosado. 2010. Use of carb counting in the dietary treatment of diabetes mellitus. *Nutricion hospitalaria* 25, 1 (2010), 18–25.

[80] Giorgia Tascini, Maria Giulia Berioli, Laura Cerquiglini, Elisa Santi, Giulia Mancini, Francesco Rogari, Giada Toni, and Susanna Esposito. 2018. Carbohydrate counting in children and adolescents with type 1 diabetes. *Nutrients* 10, 1 (2018), 109.

[81] Edison Thomaz, Irfan Essa, and Gregory D Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1029–1040.

[82] Edison Thomaz, Aman Parnami, Irfan Essa, and Gregory D Abowd. 2013. Feasibility of identifying eating moments from first-person images leveraging human computation. In *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference*. 26–33.

[83] Niels Van Berkel, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2017. Gamification of mobile experience sampling improves data quality and quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[84] Brian Wansink and Sibylle Kranz. 2013. Who's using MyPlate? *Journal of nutrition education and behavior* 45, 6 (2013), 728–732.

[85] Miranda Westfall, Sarah E Roth, Monique Gill, Alec M Chan-Golston, Lindsay N Rice, Catherine M Crespi, and Michael L Prelip. 2020. Exploring the relationship between MyPlate knowledge, perceived diet quality, and healthy eating behaviors among adolescents. *American Journal of Health Promotion* 34, 7 (2020), 713–721.

[86] Weiyu Zhang, Qian Yu, Behjat Siddiquie, Ajay Divakaran, and Harpreet Sawhney. 2015. "Snap-n-Eat" Food Recognition and Nutrition Estimation on a Smartphone. *Journal of diabetes science and technology* 9, 3 (2015), 525–533.

[87] Fengqing Zhu, Marc Bosch, Insoo Woo, SungYe Kim, Carol J Boushey, David S Ebert, and Edward J Delp. 2010. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE journal of selected topics in signal processing* 4, 4 (2010), 756–766.

Received January 2022; revised April 2022; accepted August 2022