MASTER

IRT-based lifestyle coaching for hypertension management

Boerhof, M.

Award date:
2015

Eindhoven, December 2015

IRT-based Lifestyle Coaching for Hypertension Management

By Boerhof, M. (Mark)

Identity number 0782532

In partial fulfilment of the requirements for the degree of

Master of Science
in Human Technology Interaction

Supervisors:
Dr. Martijn Willemsen (TU/e)
Prof. dr. Wijnand IJsselsteijn (TU/e)
MSc. Mustafa Radha (TU/e & Philips Research)

Keywords: Hypertension Management; Item Response Theory; Rasch; Lifestyle Coaching; diet; sodium-intake; exercise

ABSTRACT

In this thesis we used Item Response Theory (IRT) to develop a scale that rank orders blood pressure related health behaviors from various domains (diet, sodium-intake and physical activity) by the degree of challenge they impose. This scale also allowed us to gauge a person's ability to apply such behaviors. We demonstrated that this model can be used to coach hypertensive patients more effectively. In an online survey around 300 adults between 40 and 60 years of age reported their engagement in lifestyle habits that had been shown to be blood pressure relevant, in the literature. These health behaviors formed a transitively ordered class of behaviors that could be used to estimate a person's ability along the scale. In a subsequent online experiment 150 hypertensives had to indicate their relative preferences for two computerized Lifestyle Coaches who either coached random, simple or tailored interventions targeted at a patient's ability level (all three conditions were contrasted with each other). According to IRT, a Lifestyle Coach without IRT-based knowledge will often recommend relatively difficult interventions, while there are also easier interventions that an individual does not perform yet. Therefore, it was expected that an IRT-based Coach should be able to coach more effectively. The results show that, indeed, hypertensives had a significantly higher intention to perform interventions from a Lifestyle Coach that provided the simplest interventions as compared to a Coach that provided relevant interventions at random (without IRT-based knowledge). As such, these results provide support for the effectiveness of IRT models and its application in health recommender systems. The results also suggest it is not required to take into account someone's ability. Tailoring interventions to a person's ability level did not perform significantly better (nor worse) than coaching at random or coaching the simplest interventions. We conclude that lifestyle coaching systems should focus on the, according to the IRT model, easiest health interventions someone does not apply yet (so-called 'low hanging fruits').

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AHA	American Heart Association
BP	Blood Pressure
DASH	Dietary Approaches to Stop Hypertension
DIF	Differential Item Functioning
DTF	Differential Test Functioning
ESC	European Society of Cardiology
ESH	European Society of Hypertension
GEE	Generalized Estimating Equations
IRT	Item Response Theory
MET	Metabolic Equivalent of Task
m-health	mobile health
MmHg	Millimeters of mercury
PCA	Principal Component Analysis
QALY	Quality-adjusted life years
VWS	Ministry of Health, Welfare and Sport

1. INTRODUCTION

Hypertension, or high blood pressure, is a key risk factor for cardiovascular diseases (WHO, 2013). Effective hypertension management is therefore a major concern for population health. Besides blood pressure medication, non-pharmacological interventions, such as regular exercise and a healthy diet, have been proven successful in hypertension management (Dickinson et al., 2006). Appropriate lifestyle modifications may not only lower or control blood pressure in hypertensive patients but also effectively delay or prevent hypertension in non-hypertensives (Mancia et al., 2013). Unfortunately, adherence to lifestyle recommendations is generally low (Uzun et al., 2009).

For most people, changing and maintaining new lifestyle habits requires guidance and support. Literature shows that health counseling is an effective measure to induce lifestyle changes and to improve long term adherence rates (Lin, O'Connor, Whitlock, & Beil, 2010). Aside from face to face counseling from a healthcare professional, current technological developments also provide computerized coaching opportunities. Mobile health technologies such as smart fitness bands and smart watches have the potential to play an important role in the future healthcare system (Mancia et al., 2013; Chiarini, Ray, Akter, Masella, & Ganz, 2013). There are, however, many challenges for developing personalized coaching systems (Mika, 2011). One such challenge is to tailor recommendations to patient needs.

Research suggests that individuals often lack the skills to develop realistic and achievable goals (Ashford, Edmunds, & French, 2010). Individuals might be unaware of all the possible lifestyle interventions, as well as their health benefits and achievability. In many psychosocial models of health behavior the achievability of a goal and its outcome expectations are regarded as very important determinants of actual behavior (Bandura, 2004). Thus, in order to enhance adherence rates it is important to help individuals to set and reach achievable and beneficial goals.

To our knowledge, there does not exist a theoretical estimation of the relative difficulty of blood pressure related lifestyle interventions. The challenge is to design effective coaching systems that integrate the knowledge of the difficulty of interventions and an individual's ability (Byrka, 2009). Recent research indicates that Item Response Theory (IRT) is a promising technique to estimate both the (relative) difficulty of health behaviors and an individual's ability to apply such behaviors (Henson, Blandon, & Cranfield, 2010; Byrka & Kaiser, 2013; Mendoza, Schram, Arcand, Henson, & L'Abbe, 2014). This knowledge allows us to construct a scale in which health behaviors are ordered by the degree of challenge they impose relative to an individual's ability. Such scale, when integrated in a coaching program for blood pressure management, may be used to set small incremental goals and to improve coaching by recommending only achievable health interventions.

Two studies have been performed to explore the use of IRT models for health recommender systems. The first study assessed whether a one-parameter IRT (Rasch) model can be used to make inferences about the relative difficulty of blood pressure related health interventions and a person's ability to apply such interventions. The second study assessed whether IRT-based lifestyle coaching improves the coaching experience of hypertensive patients. Participants had to indicate their preference for three computerized Lifestyle Coaches who coached subsets of lifestyle advice with varying difficulty. Results from this thesis suggest that IRT-based lifestyle coaching is more effective than coaching without such models.

2. LITERATURE REVIEW

Hypertension is one of the most prominent risk factors for cardiovascular diseases, the leading cause of death worldwide (WHO, 2013). Cardiovascular complications of hypertension account for approximately 9.4 out of the 17 million deaths from cardiovascular disease a year. It is also one of the most common chronic disorders in the Netherlands (Blokstra et al., 2011). About a third of the Dutch population between 30 and 70 years of age is hypertensive. Whether an individual is diagnosed with hypertension depends on blood pressure readings which consist of two numbers; a systolic (upper) value which measures the pressure in the arteries when the heart beats, and a diastolic (lower) value which measures the pressure in the arteries between heart beats. Hypertension is defined as a blood pressure above 140 mmHg systolic or above 90 mmHg diastolic. Even pre-hypertensive blood pressure levels, i.e. above 120/80 mmHg but below 140/90 mmHg, have been found to increase the risk of stroke (Lee et al., 2011).

In 95% of the cases there is no clearly identifiable cause of hypertension. This is labeled essential or primary hypertension. The remaining 5% of the people have a clearly identifiable cause, such as a kidney disease or drug side effects, commonly known as secondary hypertension (Beevers, Lip, & O'Brien, 2001). Important risk factors for developing primary hypertension are genetic, demographic and behavioral factors. Little is known about the contribution of genetic factors, but the implications of demographic and behavioral factors are well characterized (Mancia, 2012). Examples of demographic risk factors are age and gender. Overall prevalence rates are higher for men than for women and as age progresses, the prevalence of hypertension increases (Figure 1). Also, on average, lower educated people have a higher blood pressure compared to higher educated people (Blokstra et al., 2011). This holds for all age categories and is independent of gender.

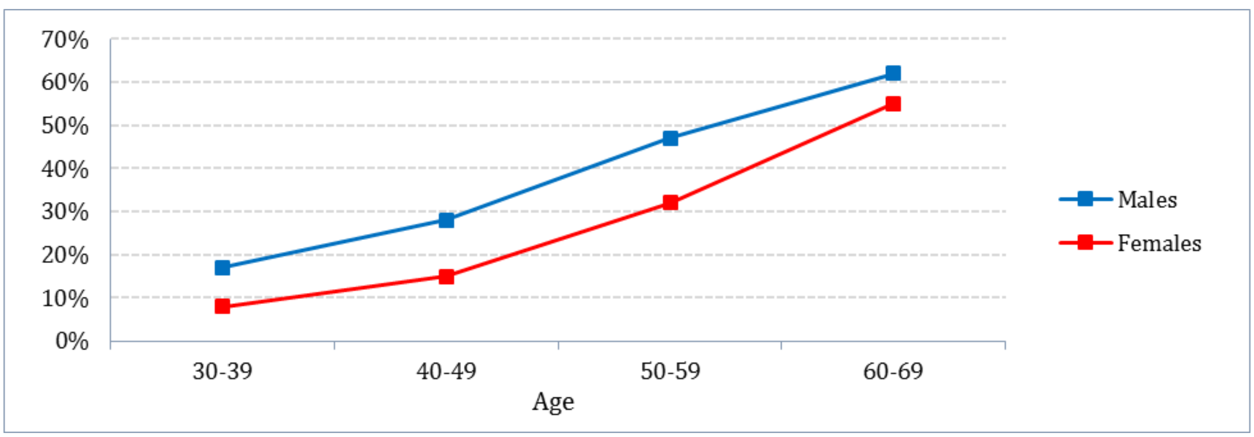


Figure 1. Prevalence rates of hypertension among Dutch adults by age and gender (Blokstra et al., 2011).

Behavioral and more controllable risk factors are lifestyle related factors such as diet, sodium-intake, physical activity and alcohol consumption (Mancia et al., 2013). Studies of health interventions show that, besides drugs, lifestyle modifications can be used to substantially lower and control blood pressure levels (Dickinson et al., 2006). In addition, lifestyle modifications might overcome the heavy economic burden and side effects of drug treatment on hypertensive patients. Despite being a prevalent health problem, hypertension is to a large extent controllable and preventable. It is, therefore, often referred to as a "chronic lifestyle disease". A major drawback, however, is the low level of adherence to lifestyle recommendations (Uzun et al., 2009).

2.1. LIFESTYLE MODIFICATIONS TO CONTROL BLOOD PRESSURE

Appropriate Lifestyle changes may safely delay or prevent the development of hypertension. Lifestyle behaviors that have been shown to reduce blood pressure are sodium reduction, physical activity, diets low in (saturated) fats and high in fruits and vegetables, maintaining a healthy weight and a moderate alcohol consumption (Table 1). Besides the positive effect on blood pressure, lifestyle changes also contribute to the control and prevention of other clinical conditions.

Table 1. Lifestyle changes to prevent and manage hypertension (modified from: Dickinson et al., 2006; Mancia et al., 2013)

Modification	Recommendation*	Impact on Systolic blood pressure (95% CI)	Impact on Diastolic blood pressure (95% CI)
Healthy diet	Dietary Approaches to Stop Hypertension (DASH)	3.4 - 8.6 mmHg	2.7 - 6.9 mmHg
Sodium reduction (1 g sodium ≈ 2.5 g salt)	Less than 2.4 g/day (does not apply to people who lose large amounts of sodium in sweat).	2.2 - 7.2 mmHg	1.8 - 3.3 mmHg
Physical activity	At least 30 minutes of moderate-intensity physical activity on at least 5 days per week.	2.1 - 10.1 mmHg	1.1 - 4.9 mmHg
Moderate alcohol consumption	Limit alcohol intake to less than 2 standard drinks per day for men, or respectively 1 standard drink per day for women.	1.4 - 6.1 mmHg	1.4 - 5.0 mmHg

* Recommended by the European Society of Hypertension (ESH) and the European Society of Cardiology (ESC)

Diet
Switching to a diet rich in fruits, vegetables, fish, nuts, low-fat dairy products and a low amount of saturated fats, such as the DASH diet, can substantially lower blood pressure (Sacks et al., 2001; Appel et al., 2006). Although a lot of Dutch adults do not meet the Dutch dietary guidelines to consume 200 grams of fruits and vegetables per day and 2 servings of fish per week (Figure 2), hypertensive patients should even be advised to consume 300-400 grams of fruits and vegetables per day (Mancia et al., 2013).

Sodium-intake
A meta-analysis based on 205 surveys shows that annually 1.65 million deaths from cardiovascular causes should be attributed to sodium consumption (Mozaffarian et al., 2014). Dutch adult consume on average 3 to 4 grams of sodium a day (Van Rossum, Buurma-Rethans, Fransen, Verkaik-Kloosterman, & Hendriksen, 2012), which is well beyond the recommended maximum of 2.4 grams per day. Recently, the American Heart Association (AHA) even reduced their sodium recommendation to no more than 1.5 grams a day, for optimal cardiovascular health (AHA [a], 2015).

Physical activity
Sufficient physical activity can lower the systolic blood pressure independent of other determinants (Whelton, Chin, Xin, & He, 2002). Although Figure 2 shows that adherence rates to physical activity recommendations are relatively high, one should keep in mind that the guidelines define a lower limit of physical activity to sustain a healthy living.

Alcohol consumption
There is a linear relationship between the prevalence of hypertension and alcohol consumption (Mancia et al., 2013). Although there is no evidence that a moderate alcohol consumption is harmful, excessive drinking is associated both with a raised blood pressure (BP) and an increased risk of stroke. Limiting alcohol consumption is an effective way to lower blood pressure (Appel et al., 2006).

Overweight
There is a strong relationship between the Body Mass Index (BMI) and hypertension, which has been well established in the literature (Kaplan, Victor, & Flynn, 2015). A 10 to 15 kg weight loss can reduce blood pressure by 4 to 20 mmHg (Chobanian et al., 2004). This makes maintaining a healthy weight a preventive measure to avoid high blood pressure. That a lot of people do not adhere to lifestyle guidelines is reflected by the high amount of people with overweight. Around 40% of the Dutch population between 31 and 69 years of age has overweight (BMI 25-≤30) and around 20% is obese (BMI≥30)(Van Rossum, Fransen, Verkaik-Kloosterman, Buurma-Rethans, & Ocké, 2011).

Other lifestyle factors
Although smoking and stress have been found to elicit short-term increases in blood pressure (Primatesta, Falaschetti, Gupta, Marmot, & Poulter, 2001; Sparrenberger et al., 2009), it is currently unclear whether these factors also contribute to long-term effects (Mancia et al., 2013).

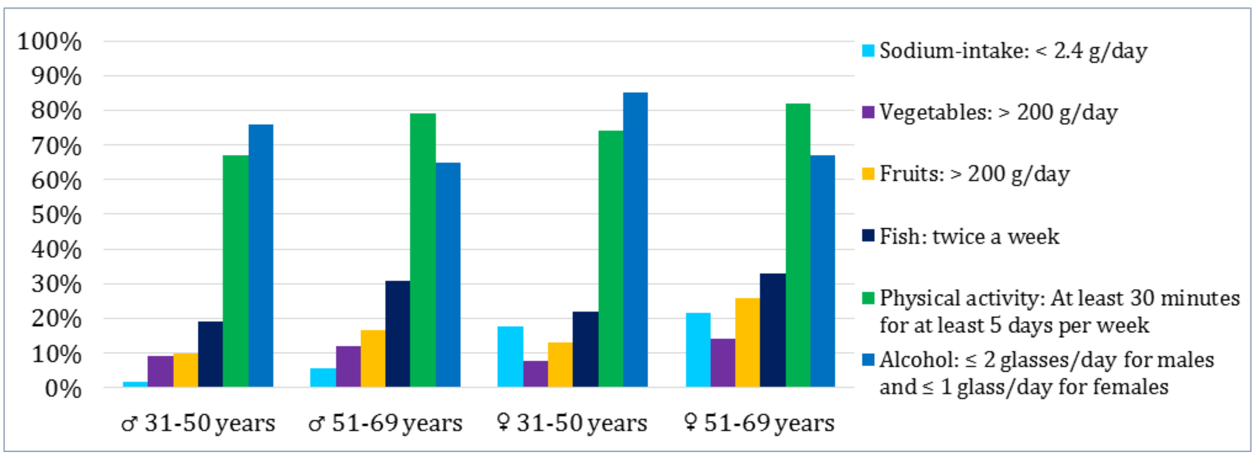


Figure 2. Adherence rates of Dutch adults to the Dutch dietary and exercise guidelines (modified from: Van Rossum et al., 2011)

2.2. Lifestyle interventions in practice

In order to prevent or control hypertension without medication, the majority of individuals has to change their lifestyle habits. Addressing risk factors such as overweight, excessive alcohol and sodium consumption and inactivity has been proven to result in an increase in quality-adjusted life years (QALY) for most people (Blair et al., 1996; Hoeymans et al., 2014). QALY is a measure of life expectancy, adjusted for the quality of life years. Lifestyle habits are, however, notoriously difficult to change on an individual basis.

2.2.1. Regulations and enforcement

It is believed that major lifestyle changes to manage hypertension have to come from societal changes (Kaplan et al., 2015). This is especially true for limiting sodium-intake. Around 80% of the sodium we consume comes from processed foods, that contains so-called 'hidden salt', of which the most important contributors are bread (26%), colt-cuts (15%) and cheese (10%)(Van Rossum et al., 2012). This makes it challenging to limit sodium consumption and to meet the sodium guidelines, even when on a healthy diet.

For over a decade, there have been nationwide lifestyle campaigns to improve population health. One such campaign called '30 minuten bewegen' aims at increasing the amount of moderate-to-vigorous intensity physical activity to at least 30 minutes per day for at least 5 days per week. Other campaigns such as 'Opzouten!' and "Grip op zout" are large scale salt-reduction initiatives that call on food manufacturers to reduce the amount of salt in their products. Research shows that, although the amount of salt in some products decreased, the average amount of salt in products did not decrease since 2007 (Consumentenbond, 2013). Although societal changes might be effective, the mentioned campaigns illustrate the difficulty of implementing them. The industry faces economic and technological challenges such as retaining food safety, flavor and preservation, which may require time, money and strict legislation to implement (Wilson, Komitopoulou, & Incles, 2012).

Despite numerous lifestyle campaigns in the past decade and the overall positive outlook with respect to general health and the life expectancy of the Dutch population (Hoeymans et al., 2014), there are still a lot of individuals that do not meet the lifestyle recommendations. As mentioned, this is reflected by the fact that (1) salt-consumption among Dutch consumers is still too high (Van Rossum et al., 2012), (2) half of the Dutch population has overweight (Blokstra et al., 2011) and (3) the amount of Dutch adults that meet the physical activity norm of 30 minutes of physical activity at least 5 days per week has remained fairly stable since 2006 (Hildebrandt, Bernaards, & Stubbe, 2013). These figures indicate that there is room for improvement.

In recent years lifestyle policies from the government were mainly focused on (media-) campaigns which stated what people should or should not do. What people want to do and are capable of doing was not sufficiently taken into account therein. Nowadays, however, the Ministry of Health, Welfare and Sport (VWS) encourages people to take more responsibility for their own health (Schippers, 2011). To facilitate this, VWS emphasizes the importance of providing reliable and accessible information tailored to information needs, accessible health facilities (e.g. healthy foods in schools) and that people experience minimal impediments to live a healthy life.

2.2.2. Lifestyle assistance

Literature indicates that, besides societal changes, health counseling is an effective measure to induce lifestyle changes and to improve long term adherence rates (Lin et al., 2010). For instance, adults with known hypertension who received lifestyle advice from a healthcare professional were more likely to report making lifestyle modifications as compared to those that did not recall receiving any advice (Viera, Kshirsagar, & Hinderliter, 2008). Hallal and Lee (2013) state that prescription of physical activity should be placed on a par with drug prescription. Also, knowledge of the risk factors of a disease has been found to be related with engagement in lifestyle changes such as healthy eating patterns, increased physical activity, and weight management (Erhardt, Roijer, Stagmo, & Uden, 2004).

It is a matter of debate whether changes should be implemented gradual or radical to have the most potential for sustainable behavior change. Although radical changes (e.g. a crash diet where someone tries to cut down in calories quickly) provide the most immediate physiological changes, they are often considered to be unhealthy and are notorious due to the high chance of falling back into old lifestyle habits. Prominent health agencies, such as the AHA, usually recommend to focus on gradual changes that can easily be adopted as lifelong habits, and state that it is not necessary to make dramatic changes all at once to manage blood pressure (AHA, 2014). This perspective is also in line with a recently introduced food concept 'Het nieuwe eten', in which The Netherlands Nutrition Centre encourages consumers to develop and maintain healthy eating habits by focusing on small gradual changes (Voedingscentrum, 2013). Implementing a simple lifestyle habit, such as using herbs and spices instead of salt during cooking, may already have a significant impact on an individual's overall daily salt consumption (Van Rossum et al., 2012). And, as little as 15 minutes of moderate intensity physical activity (e.g. walking) can reduce all-cause mortality (Wen et al., 2011). This suggests that even the smallest lifestyle changes can reduce the risk of health issues (Hill, 2009).

Aside from face to face counseling from a healthcare professional, current technological developments also provide computerized coaching opportunities. Mobile health technologies such as smart fitness bands and smart watches have the potential to play an important role in the future healthcare system (Mancia et al., 2013; Chiarini et al., 2013). Advances in ubiquitous computing opens up opportunities for integrating personalized coaching systems in e-health applications. There are, however, many challenges for developing successful coaching systems (Mika, 2011). One such challenge is to personalize health recommendations. Optimizing the personalization of health recommender systems and coaching programs has the potential to enhance patient engagement and to improve adherence rates to lifestyle changes.

2.3. Tailored lifestyle coaching

For most people changing and maintaining new lifestyle habits requires guidance and support. Ashford et al. (2010) showed in a meta-analysis of intervention studies that when an individual's goal is set by an interventionist their self-efficacy, i.e. their belief in their own competence to accomplish a behavior, significantly increased. This was not the case when individuals set their own goals. They argue that their findings indicate that most individuals lack the skills to develop realistic and achievable goals without appropriate guidance.

In many psychosocial models of health behavior self-efficacy is regarded as one of the most prominent determinants of health behavior (Bandura, 2004). According to Social Cognitive Theory (Bandura, 2004) self-efficacy influences behavior both directly and indirectly through other constructs such as goals (motivations). The relationship between goals and self-efficacy is believed to be bi-directional. Individuals who have a high self-efficacy tend to set more challenging and ambitious goals, whereas achievable goals have a positive influence on an individual's perceived self-efficacy. In order to enhance self-efficacy and to promote actual behavior change it is important to help individuals to set and reach achievable goals.

People often have abstract general health goals in mind such as "I want to have a normal blood pressure", "I want to lose weight" or "I want to eat healthier". In order to pursue these health goals it is important that individuals have a concrete idea of specific things they can do to change their behavior. For instance, to make it more practical, the AHA provides a framework on how to implement diet and lifestyle recommendations (Lichtenstein, 2006). An example of such recommendation is to use liquid vegetable oils in place of solid fats. Such practical guidelines can be used to aid health practitioners and m-health app developers in giving appropriate lifestyle advice. These smaller, more proximal, goals that contribute to the long-term goals are similar to the behavioral intention construct in the Theory of Planned Behavior (one of the most commonly used theories that links beliefs and behaviors), which indicates an individual's readiness to perform a certain behavior (Bandura, 2004).

As mentioned a key element for sustainable behavior change is that objectives should be realistic and achievable. Whether a recommendation meets these criteria largely depends on the difficulty of the proposed intervention and an individual's competence. For example, an active person who exercises four times per week might be willing to increase his or her exercises to five times per week, but a less active person might rather start with avoiding fast food or a walk around the block. Since it is often recommended to implement small changes gradually and there is a whole list of practical recommendations people can do to improve their health, it is not always straightforward which things to recommend first. As such, it is important to understand the ability of an individual and to take into account the difficulty of recommendations in order to develop effective intervention strategies. Unrealistic and difficult advice will most likely not lead to sustainable behavior change.

Besides achievability, the perceived benefits (positive outcome expectations) of an intervention is also a prominent predictor of behavior change (Bandura, 2004), especially when it comes to health-related behaviors. When people have been diagnosed with a disease such as hypertension their urge to change their behavior generally gets higher, which has a positive influence on the perceived benefit of health interventions. Unfortunately, interventions with the highest benefit often come at the highest cost. For instance, 30 minutes of physical activity per day would have a higher health benefit than 15 minutes of physical activity per day, but is also more demanding. This suggests that individuals have to make a trade-off between achievability and health benefit. An intervention that has an optimal balance between achievability and health benefit is most likely the most appropriate recommendation.

2.4. MODELLING LIFESTYLE BEHAVIORS

Unhealthy lifestyle habits such as excessive fat-, sodium- and alcohol-intake and a low amount of physical activity often occur in combination (Schippers, 2011; Héroux et al., 2012). Therefore, ideally, preventative interventions are targeted at multiple causes of chronic lifestyle conditions (Svetkey et al., 2003). Health promoting behaviors are, however, rarely seen as a single homogeneous class of behaviors and as a consequence these behaviors are often treated health domain specific. Most campaigns and coaching programs focus on one specific aspect of healthy living such as fat consumption, sodium-intake or physical activity.

Using Item Response Theory (IRT) Byrka and Kaiser (2013) demonstrated that different categories of general health behaviors (sustenance, hygiene, stress recovery, risk prevention and physical exercise) can be combined on a single scale to form a transitively ordered class of behaviors. These results suggest it might be possible to combine blood pressure related interventions from various health domains on one scale. This offers opportunities to rank the difficulty of behaviors within and between intervention categories. However, Byrka and Kaiser did not find a strong connection between dietary behaviors and physical exercise, i.e. there was not a very consistent item ordering for these behaviors when modelled on one scale. They suggest this might come due to the fact that exercise behaviors were perceived as more difficult as a whole, which caused them to tap a different part of the underlying trait. Also, their model contained only one behavior related to sodium-intake. Although Mendoza et al. (2014) used IRT to model a whole set of behaviors related to sodium-intake, they only addressed sodium restriction and not the full range of behaviors that have a positive influence on blood pressure.

The current study will extend this research by developing a scale that encompasses a large number of lifestyle behaviors from various categories (diet, sodium-intake and exercise), that have been proven to have a positive influence on blood pressure. This scale may help to determine which interventions have the most potential for sustainable behavior change and be used to coach only achievable interventions in a lifestyle coaching program.

2.4.1. Scale construction

In scale construction and evaluation, both factor analysis and IRT are frequently used for measuring attitudinal and behavioral constructs (Sick, 2011; Dima et al., 2014). A key difference between these two models is that IRT analysis specifically aims at measuring a single construct whereas factor analysis is most commonly used to expose multiple traits. IRT modelling is theory based, i.e. it should be hypothesized that the selected items form a single dimension, and is not designed for exploratory dimensionality analysis such as exploratory factor analysis. If items are hypothesized to measure multiple sources of variance these items must be subdivided a priori and a separate IRT analysis has to be conducted for each subscale.

In addition, factor analysis is a correlational model, i.e. highly correlating items measure a single construct. This implies that when an individual engages in one item of a construct (e.g. diet) it is very likely that this person also engages in the other items of the construct. It has been argued that large differences in terms of item difficulties can be problematic for factor analysis (Sick, 2011). An easy item may not correlate strongly with a difficult item, even if these items are designed to tap the same dimension. This can be an issue for health behavior modelling because of the wide spread in adherence rates (see Figure 2).

IRT on the other hand functions best when items show a wide spread in difficulty. This is due to the fact that IRT models are probabilistic and hierarchical (they allow for item ordering). According to this theory the likelihood of engaging in an item is influenced by one's ability and item difficulty. A large difference in terms of item difficulties ensures proper item and person discrimination. It is expected that difficult items are only attained by the most able persons. This directional relationship, which is inherent to IRT models and absent in factor analysis, makes it particularly suitable for analyzing items with varying difficulties that represent the same underlying trait (Dima et al., 2014).

2.4.2. IRT models

IRT models are commonly used to measure a student's mathematical ability and to order items in terms of their difficulty. The items are, however, not restricted to multiple choice questions for math tests, but can be any kind of ordered observation, such as rating or Likert scales. Recent research indicates that IRT is also suitable for modelling health-related behaviors (Henson et al., 2010; Byrka & Kaiser, 2013; Mendoza et al., 2014; Kleppe, Lacroix, Ham, & Midden, 2015). Such scales provide useful insight in the relative difficulty of health behaviors and, analogous to measuring mathematical ability, it may be used to measure someone's engagement in hypertension management. IRT models rank-order items by execution rates. In other words, an intervention is considered to be easy if almost everyone performs it. For this to be successful, it is important that individuals are aware of an intervention. For instance, it might be that almost no one performs an unfamiliar intervention, but that they actually like it due to its novelty. This would violate the difficulty/achievability relationship.

The application of such models to health behaviors is based on the idea that individuals choose to engage in behaviors that are in line with their attitudinal goals (Kaiser, Byrka, & Hartig, 2010). Individuals who do everything by bike, avoid sodium, follow a DASH diet and exercise 5 times a week will very likely be highly committed to staying or getting healthy. In other words, individuals who have a high engagement in hypertension management will most likely execute more demanding health behaviors than those who are not. The assumption is that there is an ordering in behavior difficulty which is comparable for the whole population. For instance, someone who exercises 5 times a week will probably not go to a local store by car, under the assumption that exercising is more difficult than going to a local store by foot or by bike. An individual who displays this behavior will be deemed inconsistent. Of course this person might have all sort of reasons to go by car; perhaps going by car is more convenient because he goes to work directly afterwards. The probabilistic nature of IRT models partly accounts for irregularities that may be imposed by an individual's preferences, characteristics or situational circumstances.

IRT contains a large family of models. The best-known models are a one-parameter (Rasch) model, a two-parameter model and a three-parameter model (Embretson & Reise, 2000). These models are named after the number of parameters used to estimate a person's ability. The Rasch[1] model is the most simplistic model which only uses a single parameter, i.e. item difficulty, to estimate the unobservable trait (an individual's ability). The two-parameter model adds an item discrimination parameter to the item difficulty parameter. An item discrimination parameter represents the sensitivity of an item for ability differences. A low item discrimination implies that most people, regardless of their ability level, have about the same probability of executing the behavior. Whereas a high item discrimination implies that a small difference in person ability has a big influence on the probability of engaging in the behavior. The three parameter model also adds a 'guessing' parameter to the item difficulty and item discrimination parameters. Guessing would be a problem for multiple choice math tests, but less for querying to what extent people engage in a certain behavior. Figure 3 displays all the discussed item parameters.

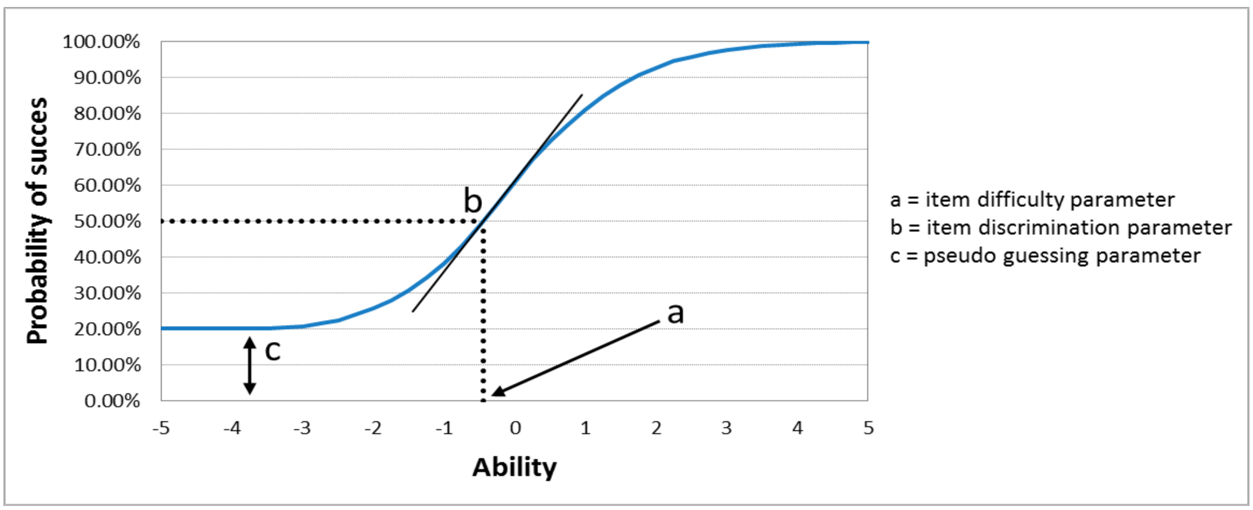


Figure 3. IRT parameters for one item characteristic curve. Item parameter a controls the item difficulty (the horizontal shift), parameter b controls the item discrimination (slope) and parameter c can be used to avoid guessing influences (the vertical shift).

The Rasch model assumes that guessing is irrelevant (noise) and that the discrimination of every item is the same. Herein item difficulty is the only item characteristic that influences person ability estimates. As a consequence, the ordering of item difficulties is the same at each ability level (Figure 4 shows that the Rasch item characteristic curves never cross). An advantage of this is that any two items can be compared independent of the person measures and that any two persons can be compared independent of the item measures (so-called specific objectivity). If the data fits the one-parameter Rasch model then it will also be possible to fit the data to the two- and three-parameter models but the reverse is not necessarily true. These properties make Rasch a simple and robust model, ideal for measurement (for an extensive description of IRT consult: Embretson & Reise, 2000).

[1] *Although Rasch is mathematically equivalent to a one-parameter IRT model and often used interchangeably, some researchers (e.g. Boone, Staver, & Yale, 2014) state that there is a substantial philosophical difference (i.e. IRT is altered to fit the data whereas in Rasch the model is superior). For the sake of understandability we also classified Rasch under the umbrella of IRT models.*

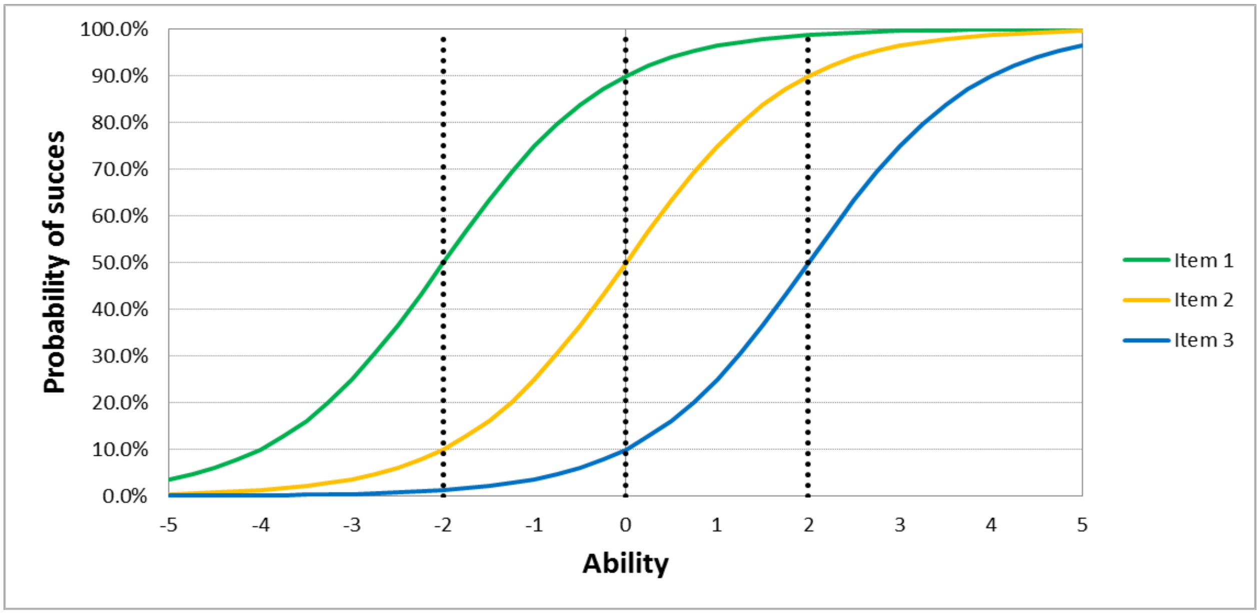


Figure 4. Rasch item characteristic curves for three items and three ability levels (dashed vertical lines). Item difficulty is from left (most easy item) to right (most difficult item). The ability level to the right is the most able person, i.e. he has the highest possibility of engaging in any of the items. The probability of engagement is always higher for item curves located to the left.

As with all IRT models, the probabilistic Rasch model holds the strong assumption that a set of items is intended to measure a single construct (e.g. engagement in hypertension management) and that the ordering of behavior difficulty is in general the same for many individuals. Eq.(1) shows the Rasch model for dichotomous data (Boone et al., 2014).

$$\theta_n - \delta_i = \ln \frac{P_{ni}}{1-P_{ni}} \qquad (1)$$

According to Rasch, the probability P_{ni} that a person (n) engages in an item (i) is determined by the logit difference between the ability of a respondent θ_n and the difficulty of an item δ_i. If a person's ability is much higher than the item difficulty, there is a high probability this person performs this behavior. A logistic relationship assures that no matter how high a person's ability is compared to the item difficulty, there is always a small chance that a person will not exert that particular behavior. In contrast, there are no certainties as well: no matter how difficult a behavior is compared to a person's ability, there is always at least a small chance that a person will perform that behavior anyway. Figure 4 illustrates the logistic relationship between item difficultly and person ability for different ability levels.

An important implication of such probabilistic relationship between the difficulty of an item and the ability of a person is that there is a 50% chance that an individual engages in items located at his or her ability level (Figure 4). From Eq. (1) it also follows that there are most likely still some relatively easy lifestyle behaviors that an individual does not perform yet. These health behaviors below one's ability level might be some of the most promising interventions to target in a lifestyle coaching program, in order to increase adherence rates to lifestyle recommendations. Unfortunately, coaching systems that recommend lifestyle interventions without IRT-based knowledge most likely randomly coach relevant interventions that an individual does not perform yet. This often causes them to miss out on the simplest interventions due to the low probability that these will be selected.

2.5. Rasch-based lifestyle solutions

Previous research primarily modelled health behaviors through Rasch to reveal the most prominent bottlenecks for public health. In these studies the difficulty of health behaviors was measured as a single latent construct, which allowed researchers to compare the relative difficulty of the modelled behaviors. These data provide novel insight that can be used to tailor educational campaigns to population needs and provide support for the need of health intervention programs.

Mendoza et al. (2014) assessed consumers' engagement in 23 sodium related interventions. These data were queried from a Canadian online food survey panel. The Rasch model showed that consumers had a poor engagement in at least 9 out of 23 of the queried recommendations, in particular behaviors related to limiting the consumption of bread and avoiding excessive salt consumption while eating in restaurants. In a similar study, Henson et al. (2010) constructed a Rasch-based scale that modelled the difficulty of healthy eating in general. They queried participants' engagement in 12 health promoting dietary recommendations related to cooking methods and consumption of specific foods. Both studies satisfied the main unidimensionality assumptions of the Rasch model, despite some scale variations based on user characteristics such as age and gender.

Research shows that Rasch models can also be used in a more practical, hands-on, fashion (Starke, 2014; Kleppe et al. 2015). For instance, Kleppe et al. (2015) developed a scale to measure patients' medication adherence more accurately. This scale was based on questions about a patient's medication-intake and provides an adherence score that represents how well a patient adheres to prescribed interventions. They were able to accurately assess adherence levels by using behaviors that covered a wide range in difficulty.

Although IRT-based lifestyle coaching is a novel and profound idea for health interventions, some related research has already been performed in the energy domain. Starke (2014) modelled energy conservation measures and used this model to gauge an individual's energy saving ability. This knowledge was then used to tailor recommendations below, at or above a person's ability level. Results showed that if participants had to rank order energy conservation measures by their appropriateness, they ordered them in correspondence with the Rasch model (i.e. the easiest items on top). These findings demonstrate the usefulness of implementing IRT models in energy recommender systems.

2.6. The present study

Both, Byrka & Kaiser's (2013) finding that health behaviors from various health domains form a single latent construct and the promising results from the energy domain provide a strong foundation for testing IRT-based lifestyle coaching for hypertension management. To our knowledge, the (relative) difficulty of blood pressure related lifestyle interventions has not been thoroughly investigated. The challenge is to design effective coaching systems that integrate the knowledge of the difficulty of interventions and an individual's ability (Byrka, 2009). IRT is a promising technique to estimate both the (relative) difficulty of lifestyle interventions (expressed as the percentage of people that engage in a certain behavior) and an individual's ability to apply such behaviors. This knowledge allows us to construct a scale in which behaviors are ordered by the degree of challenge they impose. In the context of gradual coaching, this scale may be used to set small incremental goals and improve personalized coaching by recommending only achievable health interventions. Furthermore, data about the extent to which the Dutch population engages in lifestyle behaviors provides valuable information that can be used to tailor public policies to population needs.

2.7. HYPOTHESES AND RESEARCH QUESTIONS

RQ1: To what extend do blood pressure interventions form a unidimensional set of behaviors that can be used to assess someone's ability to make lifestyle changes?

Based on previous Rasch-based research that successfully modelled health behaviors, the current research tests if it is possible to construct a unidimensional scale that orders lifestyle interventions by the degree of challenge they impose and that can be used to assess someone's ability to make blood pressure related lifestyle changes, which results in the following hypothesis:

H1: BP-related lifestyle behaviors form a unidimensional scale that can be used for coaching.

To test this hypothesis, we will model BP-related lifestyle behaviors from various health domains by employing a Rasch analysis and testing the goodness of fit. Additionally, we will perform a Principal Component Analysis (PCA) on the residuals to examine whether all unexplained variance is noise.

Based on previous research it is also expected that although some subgroups of individuals will differ in their engagement in specific interventions, the general ordering of the developed scale will be the same for these groups. For instance, males might find it more difficult to avoid red meats than females, as found by Mendoza et al (2014). Also, hypertensives and people that received lifestyle advice from a healthcare professional might apply more interventions related to sodium reduction than non-hypertensives and people that did not receive any advice. They are considered to be more aware about their own sodium consumption and to have a higher perceived risk. These expectations lead to the following hypothesis:

H2: Although subgroups differ in their engagement in some behaviors, their scales are highly similar.

In an attempt to test this hypothesis we will perform Differential Item Functioning (DIF) for Gender, education, lifestyle advice and blood pressure. This analysis contrast two independent subscales of item measures with each other to expose differences in item locations. Additionally, Differential test Functioning (DTF) will be applied to investigate whether the calculated ability scores and item ordering are influenced by the item invariance.

We also expect that some subgroups of the population differ in their engagement in hypertension management. For instance, people with a non-optimal health status, such as individuals that have been diagnosed with hypertension, will most likely engage less in healthy lifestyle habits. Because knowledge is known to have a positive influence on engagement in health behaviors, we expect a relatively high level of engagement for people with a higher level of education and/or people who received lifestyle advice. Also females are expected to have a higher ability on the constructed scale than males. Furthermore, it is reasonable to assume that the general population is aware of the implications of excessive salt consumption. However, due to the amounts of 'hidden salt' in our foods it is very hard for individuals to track the amount of salt that is consumed on a daily basis. This may cause people to underestimate their actual salt consumption, up until the point that they are diagnosed with hypertension. Therefore, we expect an overall lower engagement in salt-related behaviors for people who did not receive any lifestyle advice and non-hypertensives.

H3: Some subgroups differ in their ability on the constructed scale

Significance tests for gender, lifestyle advice, BP and education will be performed to evaluate whether there are substantial ability differences based on relevant user characteristics.

RQ2: Does IRT-based lifestyle coaching increase one's intention to perform blood pressure related health interventions?

Recent research from the energy domain indicates that people prefer recommendations below their ability. A lifestyle coaching systems without IRT-based knowledge would coach relevant interventions, for which individuals do not meet the guidelines, at random. Due to the probabilistic nature of IRT-models such coaching systems will most frequently select relatively difficult interventions (i.e. above an individual's ability level, as described in paragraph 2.4.2), while there might also be relevant 'easy' interventions that an individual does not perform yet. Targeting these interventions more effectively potentially increases adherence rates to lifestyle recommendations.

Whether individuals perceive health interventions below their ability always as the most appropriate recommendations is debatable. For instance, would individuals also opt for the easiest interventions if an alternative set of interventions clearly has a higher health benefit? Unfortunately, interventions with a high health benefit often come at a high cost (e.g. 30 minutes of exercise has a higher health benefit than 15 minutes of exercise, but is also more demanding). Based on this rationale this research hypothesizes that IRT models rank order blood pressure related health interventions not only on achievability but also on health benefit (in an inverse relationship):

H4: The achievability of advice is negatively correlated with health benefit.

To test this hypothesis we will annotate all health behaviors with their estimated health benefit and correlate these estimates with the item ordering of the Rasch scale. Because of this expected relationship, it is also hypothesized that interventions around an individual's ability level provide an optimal trade-off between the achievability and health benefit of interventions, both of which are important predictors of actual health behavior. In sum, based on previous research from the energy domain, we expect that a Lifestyle Coach that presents the simplest advice (so-called 'low-hanging fruits') outperforms a Coach that presents advice at random, which is most representative for coaching without IRT-based knowledge (baseline condition). And, that a Coach that tailors advice will perform even better. This leads to our final hypothesis:

H5: Tailored advice > Simple advice > Random advice.

In an attempt to test H5 we will evaluate all three coaching strategies, in pairwise comparisons, on a participant's intention to perform the presented lifestyle advice. From this, significance tests will be performed between all conditions to determine which Coach has the most potential for sustainable behavior change.

To provide more insight in the factors at play behind an individual's intention we will also measure engagement, recommendation quality, perceived health benefit, perceived achievability and the (objective) relative difficulty between the two presented sets of lifestyle advices. These factors will be examined in a more exploratory fashion. In line with H4, it is expected that hypertensive patients have a higher perceived health benefit for difficult advice and that they perceive easy advice as more achievable, in correspondence with the Rasch model. It is also expected that the recommendation quality, engagement and relative difficulty will be predictors of the intention to follow certain advice. A mediated path analysis will be performed to investigate which of these factors predict intention and whether the objectively measured difficulty is also a predictor of achievability and health benefit.

3. STUDY 1: SCALE CONSTRUCTION

In this thesis two studies have been performed to explore the use of IRT modelling for health recommender systems. In the first study we aimed to provide an answer for RQ1, i.e. whether the Rasch model can be used to make inferences about the relative difficulty of blood pressure related health behaviors and a person's ability to apply such behaviors.

3.1. METHOD

3.1.1. PARTICIPANTS

Data was acquired from an online survey for which participants were recruited by a recruitment agency (PanelClix). Participant received 100 clix (equivalent to 1 euro) for finishing the survey. The study was reviewed and approved by the internal ethics committee of Royal Philips. 319 participants completed the survey. Panel members were included if they were living in the Netherlands and between 40 and 60 years old and were asked, in a screening question, which type of disease they had (if any). If they selected to have an elevated blood pressure they were considered to be hypertensive, else they were considered to be normotensive and/or unaware of having hypertension. Quota sampling was applied to ensure that half of the respondents were hypertensive and the other half of the respondents were normotensive and/or unaware of having hypertension.

A small preliminary investigation to determine the validity of responses showed that 28 participants completed the survey within 5 minutes, which was unrealistic given that pre-tests indicated a completion time between 8 and 14 minutes. Another 6 participants structurally gave the same answer to series of consecutive questions. It is unlikely that thoughtful respondents would display such response behavior. These 34 highly suspicious respondents were removed from the dataset. 78 (27%) out of the remaining 285 respondents who indicated to cook never or only sometimes were subsequently excluded from 5 dietary behaviors related to cooking (their responses to behaviors 17, 19, 35, 36 and 42 in appendix A were changed to 'not applicable'). One participant that indicated to be in a wheelchair was excluded from all exercise behaviors.

Table 2. Demographic characteristics of participants with and without hypertension

Characteristics	(pre-) hypertensive (n = 142) (%)	Non-hypertensive (n = 143) (%)	Total (n=285) (%)
Gender			
Male	43.7	53.5	48.6
Female	56.3	46.5	51.4
Age			
40-49	30.8	61.8	50.3
50-59	69.2	38.2	49.7
Education			
Less than high school	2.1	1.4	1.7
High school	22.5	18.1	20.3
Secondary vocational education	45.1	52.1	48.6
Bachelor's degree	23.2	22.2	22.7
Master's degree and above	7	5.6	6.3
Other	-	0.7	0.3

Table 3. Health-related characteristics of participants with and without hypertension

Characteristics	(pre-) hypertensive (n = 142) (%)	Non-hypertensive (n = 143) (%)	Total (n=285) (%)
BMI			
< 18.5 (underweight)	-	0.7	0.4
≥ 18.5 and < 25 (normal weight)	30.5	41.1	35.8
≥ 25 and < 30 (overweight)	36.2	36.9	36.5
≥ 30 (obese)	33.3	21.3	27.3
Lifestyle advice from a healthcare professional			
Diet advice	27.5	6.9	17.1
Sodium advice	26.1	2.1	13.9
Exercise adive	16.2	6.3	11.2
Diet, sodium and/or exercise advice	50	11.8	30.8
Other			
Smoking (daily or almost daily)	24.6	34	29.4
Multiple chronic lifestyle conditions	48.6	17.4	32.9
Physical disabilities	31	14.6	22.7

Out of the 285 participants, 142 (50%) had an elevated blood pressure, 120 (42%) were normotensive and 23 (8%) did not know their status. From the 142 participants that had an elevated blood pressure 32 (23%) were pre-hypertensive and 110 (77%) were hypertensive of which 71 (50%) had ever received doctor's advice related to diet, physical activity or sodium-intake and 126 (89%) were on medication to control blood pressure.

3.1.2. Procedure

Participants conducted an online survey which took about 8-14 minutes to complete. First, they received a short introduction about the study, for which they had to give their consent. Next, participants had to indicate their engagement in 63 lifestyle interventions. Most of the interventions were retrieved from international guidelines for hypertension management (Chobanian et al., 2004; Lichtenstein, 2006; Mancia et al., 2013; Kaplan et al., 2015). These interventions were subdivided over 3 categories of 21 items; diet, sodium restriction and physical activity. The categories were presented in random order. Balancing the number of items between intervention categories is a preventive measure to neutralize the impact of potential sub-dimensions (Linacre, 2015). To ensure proper unidimensionality tests the difficulty of interventions was also balanced between categories. Most of the difficulty assumptions were based on previous research that modelled the difficulty of these health behaviors (Henson et al., 2010; Byrka & Kaiser, 2013; Mendoza et al., 2014). Finally, participants answered some demographic and health-related questions, and were thanked for participating in the survey and redirected to the PanelClix website. The content of the questions was as follows:

- **Physical activity**: Engagement in exercise recommendations was measured by querying a participant's daily or weekly amount of physical activity. More specific behaviors were queried using a dichotomous (yes/no) or a 4 point interval scale (Rarely/Never, Sometimes, Often, Always). For example, dichotomous for *'I hold a membership for sport facilities'* and a 4 point frequency scale for *'I stand during phone calls'*.
- **Diet**: The daily or weekly intake was queried for a subset of DASH items to check adherence to serving recommendations (see Kaplan et al., 2015 for more information). A second set of

dietary behaviors was queried to gain insight into more specific behaviors, such as limiting the consumption of sugary beverages or red meats. Engagement in these behaviors was measured on a 4-point frequency-scale (Rarely/Never, Sometimes, Often, Always).

- **Sodium restriction**: Questions related to sodium-intake were extracted from Mendoza et al. (2014) who assessed engagement in sodium-intake behaviors in Canada. The list of behaviors was based on expert opinions and national (Canadian) studies and surveys on sodium. All items were measured on a 4 point frequency scale (Rarely/Never, Sometimes, Often, Always).
- **Demographics**: Demographic questions were queried to qualify the target population under study. These questions were about age, weight, height, gender and education.
- **Health status**: This included questions on smoking, physical limitations, dietary limitations, medication and chronic lifestyle conditions (diabetes, obesity, cardiovascular disease). These complications may mediate a person's ability to adopt certain lifestyle patterns. Participants were also asked whether they are diagnosed with hypertension (treated or untreated) and if they had ever received lifestyle advice from a healthcare professional.

3.1.3. Analysis

A Rasch analysis was performed to estimate the extent to which subjects endorsed each of the 63 lifestyle interventions and to develop a scale that transitively orders BP-related lifestyle interventions in terms of their difficulty. All survey scales were transformed to dichotomous data. For the 4-point frequency scale the cut-off was between sometimes and often (0 = Rarely/Never and Sometimes, 1 = Often and Always). The cut-off for the dietary and physical activity recommendations was based on the guidelines (e.g. if participants indicated to eat less than 4-5 servings of vegetables this was coded as a 0, otherwise 1). DASH items were recoded into binary variables according to the recommended amount of servings which was based on a participant's calorie needs (see Appendix B for the calculations). Responses to the behaviors '*I add salt at the table*', '*I add salt during cooking*' and '*I consume more than 2 drinks per day (1 for women) or more than 14 drinks per week (8 for women)*' were inverted. Winsteps version 3.91.0 was used for analysis (Linacre, 2015).

3.2. Results

3.2.1. Descriptive analysis

Winsteps provides internal reliability statistics by calculating item and person separation scores[2]. Person and item reliability statistics range between 0 and 1 and reflect the spread of the scale and the consistency of person and item ordering. The model had an item reliability of 0.99 and a person reliability of 0.84, which indicates that 3 different levels of person measures can be identified. These reliability measures are analogous to Cronbach's alpha (Linacre, 2015) and indicate a stable scale ($\alpha > 0.8$) with a sufficient item and person spread. A wright map in appendix C shows that the average person ability is slightly lower than the average item difficulty, which implies that most persons engage in less than 50% of the presented health behaviors. The Wright map also shows that the persons and items were distributed like a Gaussian along the scale, in correspondence with the expected population and item distribution. These preliminary analyses indicate that for most persons there are enough intervention at and under their ability level. This is promising for IRT-based lifestyle coaching in which we aim to target these interventions.

[2] *Winsteps provides two reliability statistics; REAL (upper limit) and MODEL (lower limit). For this research the REAL reliability statistic was used as it is more conservative.*

3.2.2. UNIDIMENSIONALITY TESTS

As hypothesized in H1, we expect the behaviors to map on a unidimensional scale. In our Rasch analysis we tested this hypothesis by assessing the goodness of fit and performing a Principal Component Analysis (PCA) on the residuals. The fit of individual items and persons can be determined by infit and outfit statistics. Person fit statistics help to identify persons that have an incongruent response pattern, while item fit statistics help to identify items for which the ordering (place in the item hierarchy) varies substantially among participants. The outfit statistic is especially sensitive to outliers. An example of 'item' outfit would be an easy item that is not performed by a number of persons with a high engagement. An example of 'person' outfit would be an individual who engages only in the most difficult behaviors. The infit statistic focusses on unexpected response patterns to items with moderate difficulty relative to a person's ability. Identifying issues of fit is generally easier by using the outfit statistic because of its sensitivity to outliers (Linacre, 2015).

Typically the outfit mean square statistic produces values with an average near 1.0, in congruence with the predefined item discrimination the Rasch model imposes. Empirically, however, items are rarely equally discriminating. The amount of the departure of a discrimination from 1.0 is an indication of the degree to which an item misfits the model (Linacre, 2015). A range between 0.5 and 1.5 suggest a reasonable item fit. Values above 1.0 indicate underfit (too much unexplained variance) and values below 1.0 indicate overfit (too deterministic), which may result in inflated reliability statistics (Boone et al., 2014).

Nineteen participants had an outfit value outside the acceptable range (< 0.5 or >1.5). Three of these participants deviated by three standard deviations or more (Z ≥ 3), which indicates that they did not fit the model well. Boone et al. (2014) suggest to investigate these poorly fitting persons. Further investigation showed no clear indications that responses given by these three participants were erroneous or invalid. To ensure that no data was removed just because it did not match the theory these persons were not removed from the dataset. Item outfit mean squares ranged from 0.57 to 1.53 and item infit mean squares ranged from 0.82 to 1.26 (appendix A). This indicates that all 63 items fitted the model sufficiently.

All three subgroups of behaviors (exercise, diet and sodium-intake) had a reasonably diverse distribution in terms of item difficulty (Figure 5). There were also no extreme deviations between the average item measures (difficulty) of exercise (M = .06, SD = 1.41), diet (M = -.23, SD = 1.70) and sodium-intake (M = .17, SD = 1.43). Both are requirements for sound unidimensionality tests.

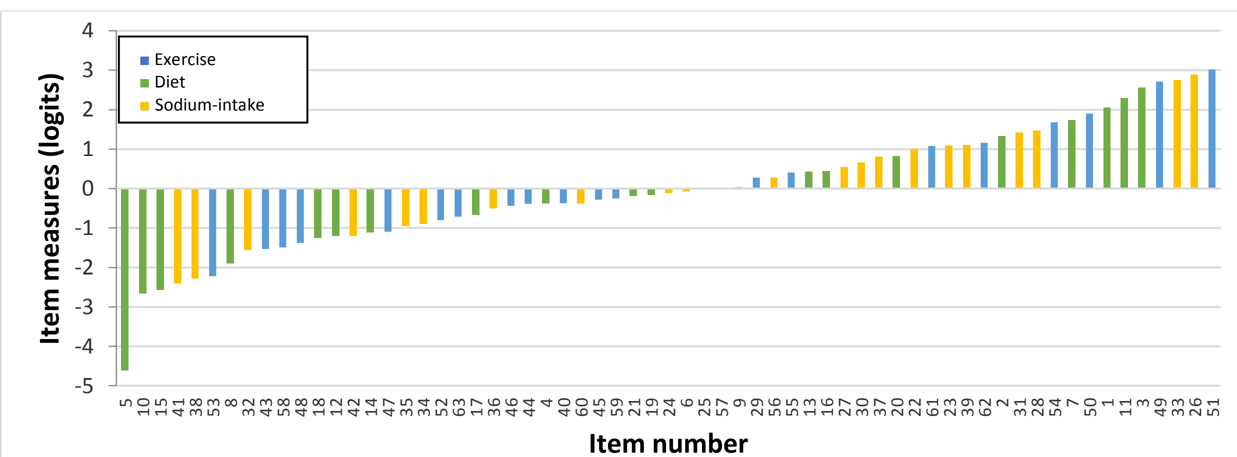


Figure 5. Item spread. Items were ordered by level of engagement, from left (high engagement) to right (low engagement).

The unidimensionality of the model was further investigated by applying a principal component analyses (PCA) on the residuals. This analysis looks for patterns in the unexplained part of the data (Linacre, 2015). The model is considered to be unidimensional if the unexplained part of the data can be attributed to noise. A PCA analysis indicated that 33.8% of the observed variance was explained by the Rasch dimension. Table 4 shows that this was equal to the expected variance (i.e. the variance if the data perfectly fitted the Rasch model and thus meets the assumption of unidimensionality). Although the explained variance was around 10 to 15 percent lower than values found in most similar studies (Henson et al., 2010; Mendoza et al., 2014) such values are not uncommon (e.g. Kleppe et al., 2015). The relatively low explained variance may partly be ascribed to the small sample size. For instance, Mendoza et al., (2014) modelled over four times as many participants, whereas Kleppe et al., (2015), who had similar results, used a sample size that was more comparable to the current study. Furthermore, Mendoza et al. modelled only salt-related behaviors whereas the current study aimed to construct a scale from three health domains, most likely reducing the explained variance.

Table 4. Standardized residual variance (in Eigenvalue units = item information units). The expected variance is the variance when the data would perfectly fit the model.

	Eigenvalues	Observed (%)	Expected (%)
Total raw variance in observations	95.2079	100	100
Raw variance explained by measures	32.2079	33.8	33.8
Raw unexplained variance (total)	63	66.2	66.2
Unexplned variance in 1st contrast	3.9274	4.1	6.2

The 66.2% unexplained variance was decomposed into 5 contrasts to examine whether it was all random noise. The first contrast explained 4.1% of the unexplained variance. This contrast was somewhat bigger than a 2.6% contrast that was obtained through a 10 simulations batch of random test data (the noise level), indicative of a potential sub-dimension. Further inspection of the first contrast showed that items which loaded high on this contrast were almost all related to exercise, whereas most of the items that loaded low on this contrast were related to sodium-intake (see appendix A for the item loadings). To investigate whether the person measures were statistically equivalent across all the items within the contrast the data was clustered into three segments of high, middle and low loadings on the first contrast. Table 5 shows that the disattenuated correlation between the upper (exercise) cluster and the lower (sodium-intake) cluster was weak (r = 0.24), which indicates that the 2 clusters measure different things.

Table 5. Approximate relationship between the person measures for the first contrast. The data was clustered into three bins according to the item loadings on the first contrast (high, middle and low).

PCA contrast	Item clusters	Disattenuated correlation
1	1 - 3	0.2462
1	1 - 2	0.7863
1	2 - 3	0.7529

Based on these indications, separating the behavioral classes should increase the explained variance. Therefore, a Rasch analysis was also performed on each class of behaviors separately (exercise, sodium-intake and diet). Although three individual models were able to explain more variance of the data, respectively 40.6% for diet, 43.4% for sodium-intake and 35.3% for exercise, the fit-statistics for the full model were better than for the separate models. Although separating the behavioral classes indeed increases the explained variance a bit, this does not outweigh the loss in item spread and the advantage of addressing multiple health domains through one scale. Furthermore, because the Rasch dimension is over eight times bigger than the second contrast, the full model is considered to be unidimensional enough to be useful for coaching purposes.

3.2.3. Item invariance

As hypothesized in H2, we expect some item variations between subgroups of the population, but that the general ordering of the items is highly similar. If the difficulty of an item varies among subgroups of test takers this is known as Differential Item Functioning (DIF). DIF performed through Rasch analysis calculates item measures for each item separately and stratifies subgroups into matching ability levels to ensure that items are only flagged if subgroups of participants have the same probability of engaging in an item. If DIF is found for many items on the scale, the final ability score might be biased. The severity of this bias can be assessed by applying Differential Test Functioning (DTF). DTF aims at validating the robustness of a scale by comparing item locations from two separate analyses.

For internal validation, the model was first tested for item invariance by performing DTF on a median split of the person measures (comparing low and high ability groups). This procedure, known as Ben Wright's challenge, tests the reliability of the scale (Bond & Fox, 2010). Items measures should remain fairly stable across the two separate analyses and ability estimates should be independent of which half of the set is used to estimate a person's ability. A cross-plot of item measures (Figure 6) indicates that not all items were experienced as equally difficult by the two subgroups, but that most items had a reasonable fit. This was confirmed by a strong correlation between the two item hierarchies (r = 0.927). Items that seemed to be more difficult for low ability persons were all about sodium considerations when buying products. Whereas items that were more difficult for high ability persons were mostly related to DASH recommendations.

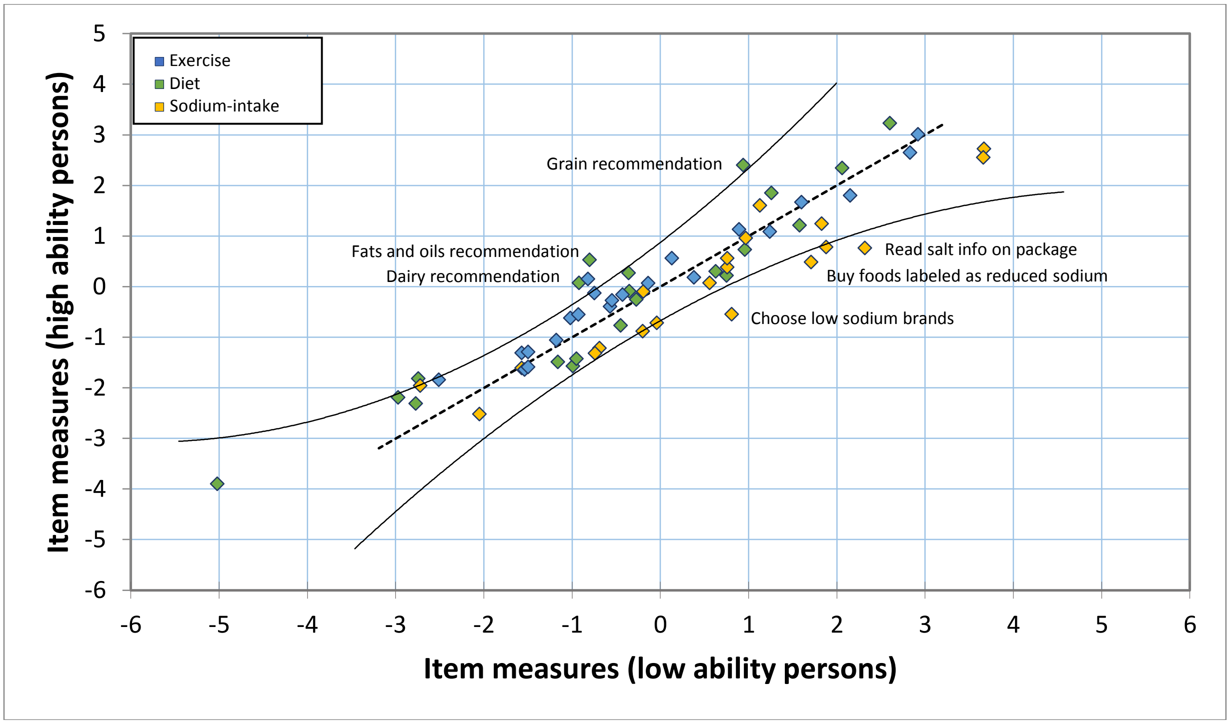


Figure 6. Cross-plot of item measures between participants with high (N=143) and low (N= 142) engagement (based on a median split). The diagonal dashed line (slope = 1) represents the line if the item measures were completely invariant. The two solid lines represent 99% confidence intervals.

Previous studies show that item invariance might also be of concern for demographic and health-related factors (Henson et al., 2010; Mendoza et al., 2014; Kleppe et al., 2015). Therefore, item invariance was further investigated for gender (male/female), level of education (low-medium/high), having received lifestyle advice from a healthcare professional (yes/no) and having hypertension (yes/no). If in a DIF test two instances of an item fell outside of the 99% confidence interval for their mean the item was flagged as potentially biased. A 99% confidence interval ensured that out of the 63 items/comparisons, on average, less than 1 item was flagged by chance. Table 6 shows the flagged items.

Table 6. Significant differences for lifestyle recommendations and tips among subgroups. Scores are logit differences compared to the base group (a negative number means that the reference group, e.g. Males, had a higher engagement).

Lifestyle recommendations	Gender		Education		Lifestyle advice		Blood pressure	
ID's	Fem.	s.e.	High	s.e.	Adv.	s.e.	HBP	s.e.
Dietary								
13	-0.09	0.26	-0.9	0.29	-0.8	0.28	-0.81	0.26
19	1.15	0.42	-0.72	0.38	0.47	0.46	0.25	0.37
Sodium-intake								
24	0.12	0.37	0.69	0.37	1.07	0.38	0.9	0.39
26	0.05	0.33	0.33	0.34	1.03	0.34	0.18	0.33
28	0.94	0.34	-0.31	0.35	0.68	0.33	0.45	0.32
31	-0.64	0.29	0.3	0.3	0.89	0.3	0.51	0.29
32	0.55	0.28	0.59	0.29	0.72	0.29	0.95	0.28
33	0.36	0.27	0.13	0.29	1.04	0.29	0.99	0.27
35	0.68	0.26	-0.25	0.28	-0.17	0.28	0.4	0.26
40	-0.2	0.27	-0.08	0.3	0.92	0.34	0.84	0.28
41	0.94	0.4	0.99	0.52	0.13	0.45	0.1	0.39
Exercise								
48	-1.25	0.33	0.25	0.34	-0.61	0.36	-0.44	0.33
51	0.15	0.27	-0.26	0.29	-0.7	0.3	-0.95	0.27
53	-0.53	0.26	-0.37	0.28	-0.82	0.29	-0.11	0.26
55	-0.72	0.27	0.11	0.29	-0.65	0.29	-0.4	0.27
56	0.09	0.27	-0.6	0.3	-0.96	0.3	-0.7	0.27
59	0	0.26	-0.34	0.29	-0.87	0.28	-0.4	0.26
60	0	0.27	-0.99	0.29	-0.63	0.3	-0.28	0.27
62	-0.42	0.28	-0.21	0.31	-0.79	0.3	-0.48	0.28
63	0	0.32	0.33	0.38	-0.89	0.33	-0.47	0.32

Highlighted numbers indicate significantly different scores ($p \leq 0.01$) between the base group and the reference group

The Rasch assumption that items should be invariant among subgroups did not hold. This was especially true for a DIF analysis between participants that did, and did not, receive lifestyle advice from a healthcare professional. Participants that received lifestyle advice had a higher engagement in 5 items relate to sodium-intake and a lower engagement in 5 items related to physical activity than participants that did not receive any advice. However, DTF analysis showed a high correlation between all of the contrasted subgroups; males and females (r = .941), low-medium and high education (r = .941), people that did, and did not, receive lifestyle advice (r = .918) and between hypertensives and non-hypertensives (r = .952)(see appendix D for the corresponding plots). These results indicate that Differential Test Functioning is unlikely and that for the great majority of items invariance can be maintained across subgroups.

In another attempt to explain more variance, separate Rasch models were constructed based on the variable that queried for lifestyle advice. A model with only participants who received lifestyle advice (N=88) did not have a significantly higher explained variance than the full model (33.8%). A model with only participants that did not receive lifestyle advice (N=197) gave marginally better results than the full model (34.9% explained variance).

3.2.4. Ability differences among subgroups

As hypothesized in H3, we expect that some subgroups of the population (e.g. based on gender, lifestyle advice, BP and education) significantly differ in their ability on the constructed scale. In order to check for ability differences among subgroups, the queried demographic and health variables were correlated with overall ability. A correlation matrix including all the queried demographic and health variables (Appendix E) showed that overall ability had a significant positive correlation with lifestyle advice (r= .209, p < .01), level of education (r = .161, p < .01), and gender (r = .182, p < .01), with women having higher scores than men.

Contrary to our expectations, overall ability did not correlate with having an elevated blood pressure (r = .028, p > .05). A correlation analysis between elevated blood pressure and ability measures calculated for three separate Rasch models (sodium–intake, diet and exercise) showed that although having an elevated blood pressure had a strong positive correlation with sodium-intake ability (r = .219, p < .01), it had a negative correlation with exercise ability (r = -.153, p < .01) and did not correlate with diet ability at all (r = .001, p > .05). A high engagement in sodium related behaviors seemed to be neutralized by a lack of physical activity.

3.2.5. Rank ordering of health benefits

As hypothesized in H4, we expect to find an inverse relationship between achievability and health benefits. In order to investigate whether the constructed scale produced such relationship, all items on the scale were ranked based on a rough estimation of their health benefits (see appendix A). Interventions were ranked within their category (diet, sodium-intake and physical activity) since they had comparable units of measurement. Sodium interventions were annotated with their impact on the daily sodium consumption. The benefits of sodium interventions were mostly extracted from Van Rossum et al., (2012), who estimated the average contribution of a whole range of food products to the daily salt consumption of Dutch individuals. Interventions related to physical activity were annotated with MET minutes. MET minutes express the time engaged in an activity related to the intensity of a task. MET minutes are calculated by multiplying the Metabolic Equivalent of a Task (1 MET is equivalent to the calorie expenditure at rest) with the average time spent on a task per day (MET values were extracted from Ainsworth et al., 2000). Dietary interventions were annotated using their expected impact on blood pressure. All DASH-related items were considered to have the maximal health impact.

Pearson correlations showed no correlation between the health benefit and achievability of dietary interventions (r = 0.038), a weak negative relationship between health benefit and achievability for sodium related interventions (r = -0.120) and a moderate to strong negative correlation between health benefit and achievability for exercise related interventions (r = -0.448).

3.3. Discussion

The descriptive analysis gives a first indication that the constructed Rasch scale can be used to coach hypertensive patients more effectively. A Wright Map in Appendix C shows a clear Gaussian like distribution for both person- and item measures, in line with the expected population and item distribution. This map also indicates that, although the majority of interventions are above a participant's ability level, for most participants there are also intervention at and under their ability. These findings underline the potential of IRT-based lifestyle coaching in which we aim to target these 'achievable' interventions.

3.3.1. Engagement in health behaviors

Perhaps not surprisingly, the most difficult dietary behaviors were related to the daily recommended intake of fruit, vegetables, grains, nuts, seeds and legumes, extracted from the DASH diet. Because the population is considered to be aware of (at least) the importance of fruit and vegetable intake, this emphasizes the importance of interventions beyond mass media campaigns.

With respect to sodium related behaviors, participants in this study engaged least in behaviors that aimed at reducing sodium-intake while eating at restaurants. Mendoza et al., (2014) also found these behaviors to be difficult for Canadians. Both in the Netherlands and in Canada, there is a need of interventions aimed at restaurant food. Probably not surprisingly, limiting the consumption of cheese was also a very difficult behavior for Dutch people (way more difficult than for people from Canada). This is of concern, as usually cheese not only contains a high amount of sodium (about 0.62 out of 100 grams), but also a high amount of (saturated) fats. Although there are types of cheese that are 'light' in fat, they still contain about 19 grams of fats per 100 grams, of which 13 grams are saturated. According to the AHA daily saturated fat consumption should be limited to less than 6% of the total calorie intake. This implies that a 2,000 calorie diet should contain less than 120 calories (or 13 grams) from saturated fat (AHA [b], 2015). Eating three 'light' cheese sandwiches a day already approaches the daily maximum. Limiting cheese consumption might therefore highly contribute to hypertension management in the Netherlands. Also few participants read sodium information on the packaging when grocery shopping (item 26 in Appendix A), especially in the low ability group. Educational campaigns may be targeted to these interventions.

Only a few participants perform exercises while brushing their teeth. It might be that most participants never thought of this intervention, or that participants find it a rather weird or useless recommendation. Also, almost no participants kept an exercise diary or tracked their daily activity. This suggests that for these interventions the health benefits do not outweigh the burden. However, rapid advancements in technological tracking devices, such as smart watches, and public acceptance of these devices make these difficulty insights time-sensitive.

3.3.2. Unidimensionality tests

The difficulty of interventions was well balanced between subcategories (diet, sodium-intake and exercise), which is a pre-requisite for sound unidimensionality tests. The fit-statistics suggested that the Rasch model satisfied the unidimensionality assumption. However, a subsequent PCA-analysis exposed a negligibly small but present second dimension (exercise behaviors loaded high on this dimension and sodium related behaviors low), which is indicative of multi-dimensionality. This finding was strengthened by a relatively weak correlation (r = 0.24) between the upper and lower clusters of the first contrast (Table 5) and is in line with results found by Byrka & Kaiser (2013). However, Byrka and Kaiser's suggestion that a low correlation between exercise and dietary behaviors might have resulted from difficulty differences between these categories was not supported. In the current study the three health domains had a comparable average difficulty.

In summary, because all items fitted the Rasch model reasonably and the Rasch dimension explained about eight times more variance than the first contrast, a potential second dimension was considered to be small enough to be ignored in further tests of the model for coaching purposes, supporting H1. Furthermore, the model was able to reliably discriminate between three different levels of person measures, indicating that participants can be grouped into three ability bins (low, medium and high).

3.3.3. Item invariance

DIF analyses for gender, education, lifestyle advice and blood pressure revealed that the invariance assumption was not empirically supported among different subpopulations. For instance, participants that received lifestyle advice structurally perceived some sodium-items as less difficult and some exercise items as more difficult than participants that did not receive any lifestyle advice. It is unclear whether merely providing advice with an IRT-based coaching system is already enough to reduce the variance or that there are serious underlying discrepancies between these two groups. Nonetheless, the hypothesis of a strictly uniform latent variable is untenable. The invariance assumption is commonly rejected in studies that modelled health behaviors through Rasch (Henson et al., 2010; Mendoza et al., 2014).

Separate IRT models based on user characteristics may reduce DIF and increase the explained variance of the total model. For the current model, subdividing the sample based on the variable '*lifestyle advice*' did not provide a substantial increase in explained variance. However, it has to be mentioned that with such small sample sizes it is often a trade-off between the purity of measurement and the spread of person measures which, together with the item spread, is one of the main parameters for the estimated explained variance.

Although some items experience DIF, this is not necessarily problematic for coaching purposes, the actual item ordering is far more important. For coaching purposes a rough ability estimate is already enough to tailor recommendations .DTF analysis between low and high ability participants demonstrated that the scale is quite robust. Subsequent DTF analyses for gender, education, lifestyle advice and blood pressure confirmed that, although some items experienced DIF, item hierarchies were highly similar between all subgroups. This suggests that the calculated ability scores are independent of subgroups and that the scale is useful for coaching purposes, in support of H2.

3.3.4. Ability differences among subgroups

As expected, females, highly educated participants and participants that received lifestyle advice engaged, on average, in more health-related behaviors than their reference groups. These findings stress the importance of lifestyle counseling. Together with the fact that for some items and subgroups the assumption of item invariance was violated, these results also highlight the importance of demographic and health-related factors on the difficulty of hypertension management.

In contrary to our expectations, no ability differences were found between people with and without have hypertension. Although hypertensives engaged more in sodium-related behaviors this effect was neutralized by their lack of physical activity. It should be noted that the high blood pressure group contained more individuals with obesity. A high BMI may make it harder to exercise. Although the causality of the effect remains uncertain, since little exercise may also cause a high BMI.

3.3.5. Rank ordering of health benefits

Given the, on average, weak negative correlation between achievability and health benefit it seems that IRT-models do not substantially rank blood pressure related health behaviors by their health benefits. Because these findings suggest that in the current scale difficult behavior is often not more beneficial than more achievable behavior we cannot accept our hypothesis (H4) that there exists a substantial inverse relationship between achievability and health benefit. This reduces the chance that our final hypothesis (H5), in which we state that hypertensive patients will prefer recommendations tailored to their ability over the simplest recommendations, holds. However, it has to be mentioned that this is merely a first indication since the ranking is based on rough estimates. Also, an individual's perceived health benefit might deviate from these findings.

4. Study 2: IRT-based Lifestyle Coaching

The first part of this study queried to what extent participants engage in various lifestyle habits, to provide an answer for RQ1. From this data we developed a scale in which lifestyle habits are ordered by the degree of challenge they impose. In the second part of the study we used this scale to provide personalized lifestyle advice. To test the effectiveness of this advice, participants had to compare subsets of lifestyle advice and to indicate their preferences. Three lifestyle coaching strategies were contrasted with each other; Lifestyle Coach 1 presented tailored advice, Coach 2 the simplest advice and Coach 3 random advice. In this study we aimed to provide an answer for RQ2: Does IRT-based lifestyle coaching increase one's intention to perform blood pressure related health behaviors?

4.1. Method

4.1.1. Participants

The survey of Study 1 was expanded with a few questions that queried a participant's preference for three different types of lifestyle coaching. As in Study 1, participants were recruited through Panelclix and received a compensation for their participation. The study was completed by 151 (pre-)hypertensives, aged between 40 and 60 years old (74 females, 74 males), whose education level ranged from primary education to university-level. The distribution of participant characteristics was comparable to the statistics provided for (pre-)hypertensives in Table 2 and Table 3 of Study 1.

A small preliminary investigation to determine the validity of responses showed that 3 participants completed the survey within 6 minutes, which is unrealistic given an average completion time between 9 and 15 minutes for this test. These respondents were removed from the analysis. Responses of the remaining 148 panel members were used for further analyses.

4.1.2. Procedure

Three conditions were constructed to evaluate the effectiveness of IRT-based lifestyle coaching and to test our final hypothesis H5. All three conditions represented a hypothetical Lifestyle Coach that presented three lifestyle advices, one from each of the three health domains; diet, sodium-intake and exercise. The manipulated factor between conditions was the difficulty of the displayed advice. Only interventions that a participant did not yet perform were presented to them. Because participants had to conduct the full questionnaire from Study 1, in which they had to indicate their engagement in all 63 health behaviors, we were able to mark and exclude all health behaviors that a participant already performed.

The Lifestyle Coach in the first condition presented interventions randomly and was considered to be most representative for lifestyle coaching without IRT-based knowledge. The two other conditions presented a Lifestyle Coach that used IRT-based knowledge in order to optimize the coaching experience. In the second condition the ordering of item difficulty that the model produces was used to present the simplest interventions. The rationale behind this condition is that individuals should have the highest probability of success for the, according to Rasch, 'simplest' interventions they do not yet perform. The third condition also included a personalization factor to present tailored interventions. In this condition, interventions were ordered according to the IRT model and only presented if they were located around a participant's ability level (see appendix C for a corresponding item map). This condition is based on the rationale that participants have to make a trade-off between the achievability of interventions and their health benefits. It was expected that coaching around someone's ability level provides an optimal trade-off between the two. However, the results of Study 1 suggest that participants only had to make a small trade-off, due to a small negative correlation between achievability and health benefit. This led us to believe that tailored lifestyle advice might be less effective than previously expected.

In the present study we were primarily interested in detecting differences among various coaching strategies. Although between-subject designs are usually realistic of real world experiences they lack a clear frame of reference, which makes them less suitable for relative judgements (Hsee & Zhang, 2010). In the present experiment we attempted to overcome this limitation by employing pairwise comparisons between two Lifestyle Coaches. This within subjects-design allows us to detect more subtle differences between coaching strategies (see appendix F for the experimental set-up).

Participants had to perform a joint evaluation between two Lifestyle Coaches to test how the difficulty of recommendations (random, simple or tailored interventions) and the ability of a person relate to the satisfaction of that person with the Lifestyle Coach. Study 1 indicated that the model was able to distinguish 3 significantly different ability levels. Based on this knowledge and more commonly used difficulty levels, in for instance gaming, we decided to split the data in 3 ability levels (low, medium, high) and three difficulty levels (easy, medium, hard). Tailoring advice implied that these difficulty and ability levels were matched. For example, easy advice was recommended to persons with a low ability. The design of the experimental set-up is illustrated in the diagram below:

Table 7. Experimental design (R = random advice, S = simplest advice, T = tailored advice).

Person ability	**Condition (X vs. Y)**		
Low	R vs. T/e	S vs. T/e	R vs. S
Medium	R vs. T/m	S vs. T/m	R vs. S
High	R vs. T/h	S vs. T/h	R vs. S

Participants got to see all three comparisons in randomized order, this partially controls for the within subjects effect. Also, because each condition was presented the same number of times a familiarity bias was not likely to be present. The exact procedure was as follows:

- **Step 1:** First, questions from Study 1 were presented. Responses to these questions were used to determine a participant's lifestyle habits and to calculate a ***user ability score***. Ability calculations were based on the percentage of lifestyle habits for which the user complied to the recommendations:

$$Ability = \frac{(Total\ responses - not\ applicable)}{correct\ responses} \qquad (2)$$

  In Eq.(4.1) a correct response implies that a person adheres to the recommendation. An analysis on the data collected in Study 1 showed that ability scores calculated through Eq.(2) correlated high (r = 0.991) with values obtained through an analysis in Winsteps, indicating a high accuracy. A participant was assigned to one of the three ability bins (low, medium high) based on the ability score. The samples in each ability bin should be equal with natural sampling as the bin borders were the empirically seen 33rd and 66th percentiles of the ability scores from Study 1.
- **Step 2**: Two hypothetical lifestyle coaches were described who either coached random, simple or tailored advice in a blind experiment, i.e. these coaches were not labeled. To ensure that DIF was not a factor at play both coaches gave three advices, one from each health category (diet,

sodium-intake and exercise). If possible, all three conditions were contrasted with each other (i.e. participants had to evaluate three comparisons). However, if there was a condition for which there were no interventions to present (e.g. if a participant already performed all of these interventions) or if two Lifestyle Coaches selected the same interventions, no comparison was displayed. Calculations based on the data from Study 1 showed that the likelihood for this to occur was acceptable and at most 30% of the times for tailored versus simple in the low ability group. The two most likely groups to present similar advice.

– **Step 3**: The user was asked to compare the two Lifestyle Coaches on the given advice based on questions related to five factors; the *Intention* to perform the recommendations for the upcoming three months, the *Achievability* of the recommendations, *Recommendation Quality*, perceived *Engagement* and perceived *Health Benefit*. *Intention* was our main factor of interest and captured with two questions. The other four factors, which were queried to investigate the factors at play in a participant's choice, were captured with three questions (see appendix F for all questions).

The survey concluded with the same demographic questions as the questionnaire that was conducted in Study 1. After the response was saved, users were redirected to PanelClix in order to evaluate the survey and to receive a compensation for completing the survey.

4.2. RESULTS

4.2.1. COACHING STRATEGIES

As hypothesized in H5, we expect to find the following ordering in terms of a participant's intention to perform the presented lifestyle advice: Tailored advice > Simple advice > Random advice. To test whether *Intention* significantly differed between the three conditions we applied significance tests for all conditions and all ability levels within each condition. A preliminary exploratory analysis suggested that the two ordinal questions that aimed to target *Intention* did not follow a normal distribution, this was confirmed by a Shapiro-Wilk test ($p < 0.001$). A Kendall's tau-b correlation showed that these questions correlated highly ($r = .789$). The two questions were reduced to one *Intention* factor since their combined reliability was also high ($\alpha = .918$). Conventional statistic books

Table 8. Relative Intention by condition and ability levels. For display purposes the 5 point Likert scale was collapsed into 3 response options, with the middle option as 'Neutral'.

Condition (*X* vs. *Y*)	**Ability**	***N***	**Difficulty (*X-Y*) [a]**	**Pick *X***	**Neutral**	**Pick *Y***	**Pick easy (Y/N)**	***p* [b]**	***p* [c]**
T vs. R	Low	39	-1.21	19	7	13	Y	0.649	0.526
	Medium	38	-0.55	19	6	13	Y	0.356	0.335
	High	51	0.41	17	14	20	Y	0.647	0.639
	Total	128	-0.31	55	27	46	Y	0.688	0.578
T vs. S	Low	39	0.01	12	14	13	-	0.683	0.785
	Medium	38	0.84	18	5	15	N	0.644	0.621
	High	51	2.17	15	10	26	Y	0.119	0.096
	Total	128	1.19	45	29	54	Y	0.438	0.419
S vs. R	Low	49	-1.21	26	8	15	Y	0.107	0.091
	Medium	45	-1.39	22	8	15	Y	0.287	0.280
	High	54	-1.76	24	12	18	Y	0.642	0.510
	Total	148	-1.50	70	28	48	Y	0.077	0.049

[a] The values represent logit differences; a positive number indicates that X was more difficult than Y

[b] One-Sample Wilcoxon Signed Rank Test

[c] One-Sample T-Test

recommend to either transform the data or to perform non-parametric significance tests if the assumptions of normality are questionable. However, research suggests that for small sample sizes parametric tests often do not perform worse than non-parametric tests, even if these data do not meet the normality assumption (Meek, Ozgur, & Dunning, 2007). Therefore, we tested the data with both a parametric T-test and a non-parametric Wilcoxon Signed-rank test. Both analyses were tested against a hypothesized value of 3 (the middle value on a 5 point Likert scale), for which a value lower than 3 indicated a preference for Coach X and a value higher than 3 a preference for Coach Y. Table 8 shows that participants had a significantly higher intention to perform simple advice as compared to random advice for both the T-test, $t(147) = -1.984$, $p = .049$, and the one-sample Wilcoxon signed-rank test (at the 0.10 level), $W = 2{,}971$, $z = -1.766$, $p = .077$. Also, most participants had a higher intention to perform simple advices as compared to tailored advices, indicating that tailoring advices based on a participant's ability level advices is not more successful than lifestyle coaching that is solely based on the item ordering. Although most participants had a higher intention to perform tailored advice than random advice, this effect was not significant. Overall, participants preferred the easiest set of recommendations, in line with the philosophy behind Rasch.

In sum, although our hypothesis H5 did not hold, participants had a significantly higher intention to perform advice from an IRT-based Lifestyle Coach that presented the simplest recommendations than from a Coach that presented recommendations at random (without IRT-based knowledge).

4.2.2. Exploratory data analysis

To provide more insight about the factors at play in a participant's choice of a Lifestyle Coach, we subjected the comparative questions to an Exploratory Factor Analyses (EFA) while controlling for the within subject effects, as each Lifestyle Coach was presented twice in three pairwise comparisons. All questions were treated as ordinal variables. The data was reduced from the four theoretically determined factors "*Achievability, Recommendation Quality, Engagement or Health Benefit*" to two factors. The first factor, which we named "*Satisfaction*", consisted of five items; A1, A2, Q3, E2 and E3 in appendix F. The reliability of this factor was high (α =.912). A second factor containted all three items from the theoretically defined construct *Health Benefit*. A reliability analysis over this set of items indicated that the reliability of this factor was also high (α = 0.941). Together, the two factors were able to explain 47.7% of the variance in the data.

Kendall tau-b correlations were performed for the three remaining constructs; *Intention, Satisfaction, Health Benefit,* and the relative difficulty (the logit difference between the two presented sets of items; X and Y) to assess the relationship between these variables. The correlations showed that *Intention* positively correlated with *Health Benefit* (r = .552), as well as *Satisfaction* (r = .741) and the relative difficulty (r=.081). A positive correlation between, for instance, *Satisfaction* and the relative difficulty implies that when Coach X gets more difficult than Coach Y (X-Y is positive), there is a higher Satisfaction for Coach Y (i.e. easy advice). *Health Benefit* also correlated with positively with *Satisfaction* (r = .611) and there was also a small but positive correlation with the relative difficulty (r = .088). *Satisfaction* significantly correlated with the relative difficulty (r = .070) at the 0.10 level.

Three separate Generalized Estimating Equations (GEE) models were used to determine whether the objectively measured difficulty was a predictor of (1) *Intention*, (2) *Satisfaction* and (3) *Health Benefit*. GEE allows us to control for the repeated measures effect. An ordinal GEE performed over the whole dataset, with the objectively measured relative difficulty (X-Y) as predictor and *Intention* as the dependent variable, showed that participants had small but significantly higher intention to perform easy advice, $\beta = .134$, $p = .034$. A GEE performed over the whole dataset, with the objectively measured relative difficulty (X-Y) as predictor and *Satisfaction* as the dependent variable, showed that easy advice did not lead to a significantly higher satisfaction, $\beta = .077$, $p = .233$. A GEE performed over the whole dataset, with the objectively measured relative difficulty (X-Y) as predictor and *Health Benefit* as the dependent variable, showed that participants also perceived a small but significantly higher perceived health benefit for easy advice, $\beta = .159$, $p = .029$.

Table 9. Participants' averaged responses to the five factors for three classes of person ability and three compared conditions (X vs. Y; T =Tailored, R = Random, S = Simple). Preference was queried on a 5 point-Likert scale (left indicates preference for X).

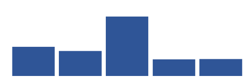
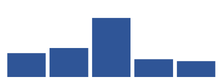
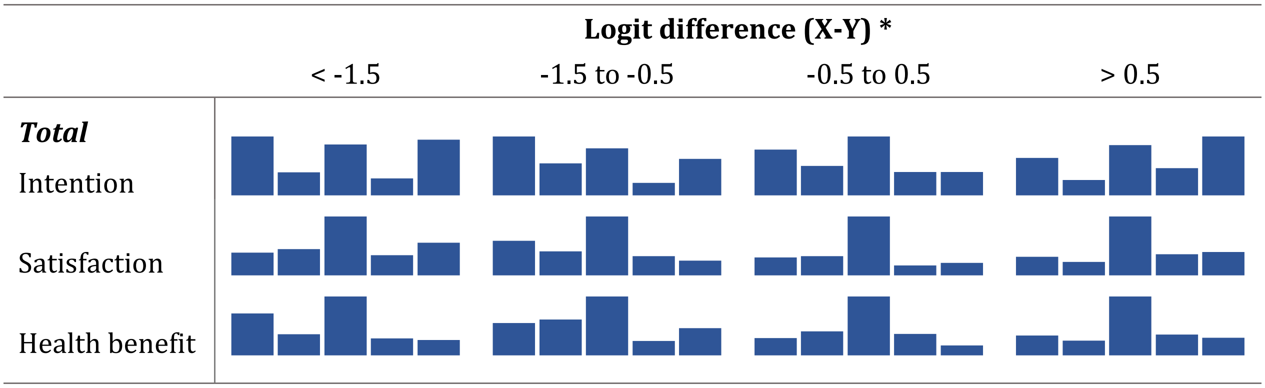
	Condition (X vs. Y)		
	Tailored vs. Random	Tailored vs. Simple	Simple vs. Random
Low ability			
Intention			
Satisfaction			
Health benefit			
Difficulty (X-Y)	(-1.33) – (-0.12) = -1.21	(-1.33) – (-1.34) = 0.01	(-1.34) – (-0.12) = -1.22
Medium ability			
Intention			
Satisfaction			
Health benefit			
Difficulty (X-Y)	(-0.31) – (0.24) = -0.55	(-0.31) – (-1.15) = 0.84	(-1.15) – (0.24) = -1.39
High ability			
Intention			
Satisfaction			
Health benefit			
Difficulty (X-Y)	(1.25) – (0.84) = 0.41	(1.25) – (-0.92) = 2.17	(-0.92) – (0.84) = -1.76
Total			
Intention			
Satisfaction			
Health benefit			
Difficulty (X-Y)	(0.10) – (0.41) = -0.31	(0.10) – (-1.10) = 1.20	(-1.10) – (0.40) = -1.50

Table 9 shows the average responses of all participants for all ability levels based on the two remaining factors. This table indicates that there was a strong tendency among participants to select the middle option, which suggests a lot of participants did not have a clear preference for one of the two Lifestyle Coaches on the underlying constructs of intention. To examine to what extend the difficulty of advice influenced a participant's response on the measured concepts we plotted a participant's response in relative difficulty bins. Table 10 indicates that participant's seemed to have a higher intention to perform the easiest set of advice and that both other constructs, *Satisfaction* and *Health Benefit*, seem to follow this trend to a smaller degree.

Table 10. Relative difference in difficulty between the two presented sets of lifestyle advice (X and Y), independent of conditions.



* A negative value implies that X (left two bars) was more difficult than Y (right two bars)

A multiple mediation path analysis was performed to investigate whether the effect of difficulty on *Intention* was mediated by *Satisfaction* and/or *Health Benefit* (Figure 7). This analysis produces estimates of the indirect (mediation) effect, the direct effect and the total (indirect + direct) effect of the relative difficulty on *Intention*. Results show that there was not a significant indirect effect of the relative difficulty on *Intention*, $\beta = .043$, $p = .242$. The direct effect of the relative difficultly on *Intention* was significant at the 0.10 level, $\beta = .032$, $p = .062$, the same holds for the total effect of the relative difficulty on *Intention*, $\beta = .075$, $p = .057$. The multiple mediation path analysis also indicates that there was not a significant direct effect of the relative difficulty on *Satisfaction*, $\beta = -.026$, $p = .637$, and not a significant effect of *Health Benefit* on *Intention*, $\beta = .026$, $p = .481$. Overall, participants had a slightly higher intention to perform easy advice and did not perceive difficult advice to be more beneficial for their health than easy advice.

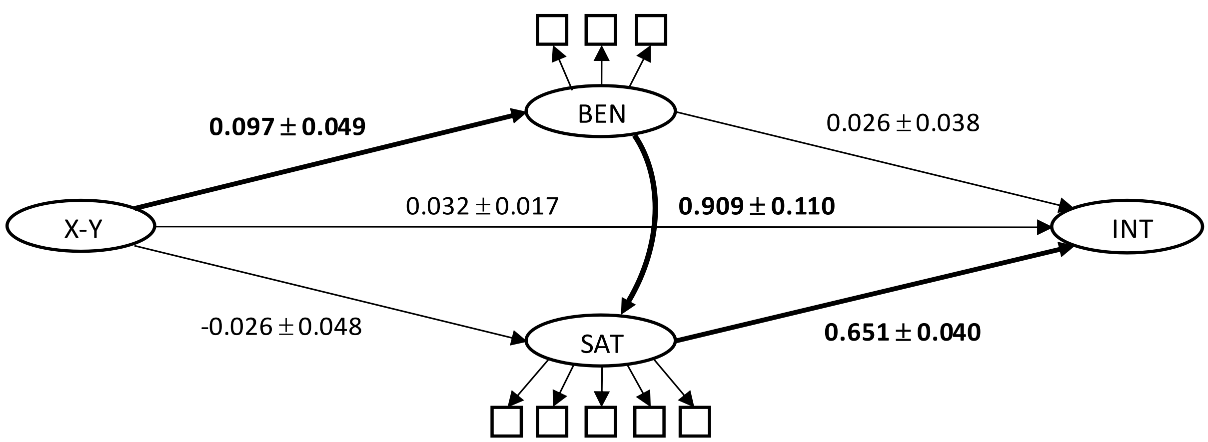


Figure 7. Multiple mediated path analysis. Intention as independent variable and the relative difficulty (X-Y), Health Benefit (BEN) and Satisfaction (SAT) as predictors. All bold paths and associated coefficients are significantly nonzero (p < 0.05).

4.3. Discussion

The results indicate that IRT-based lifestyle coaching performs significantly better than coaching at random. The results also suggest that this is caused by the fact that random advice often implies difficult advice. Participants had a significant higher intention to perform recommendations from a Lifestyle Coach that provided simple advice as compared to a Lifestyle Coach that provided random advice. It should be noticed that these were also the conditions with the biggest difference in item difficulties (an average difference of 1.5 logits). Overall, participants preferred the easiest set of recommendations, in congruence with the philosophy behind the Rasch model and findings from the energy domain (Starke, 2014). As such, this study validates the usefulness of health behavior modelling through IRT and provides support for its application in lifestyle coaching.

Our results also show that hypertensives did not have a significantly higher intention to follow tailored advice as compared to random or simple advice. This suggests that it is not necessary to take into account an individual's ability level and that lifestyle coaching should focus on the easiest interventions someone does not perform yet. The fact that tailored advice was not significantly better (nor worse) might be explained by three main reasons: First, although we found a small negative correlation between the achievability of advice and their health benefits, participants did not perceive difficult advice to be more beneficial for their health, indicating that they did not have to make a substantial trade-off. Second, the perceived health benefit also did not seem to be a direct predictor of the intention to perform recommendations. And third, tailored advice may not have provided an optimal trade-off between achievability and health benefit (or not in all situations). For instance, tailored advice for high ability person's, i.e. the most challenging set of interventions, might even have been too hard for these individuals.

Although Lifestyle Coaching based on the item ordering of the Rasch model performed significantly better than random, it has to be mentioned that there were still a lot of participants who either did not have a clear preference or who preferred a Lifestyle Coach that presented random advice (as shown in Table 9). These results are explained by the fact that (1) random recommendations will sometimes have about the same difficulty as the easiest recommendations, making it impossible to choose between the presented sets of advice in terms of difficulty. And, (2) due to the probabilistic nature of IRT-based lifestyle coaching and the wide variety of individual preferences and circumstances, there is always a chance that IRT-based lifestyle advice does not match a person's situation or preferences.

The fact that *Intention*, *Satisfaction* and *Health Benefit* all followed a similar pattern might have resulted from the experimental design. All theoretical constructs with their associated questions were presented in a single question block (see Appendix F). Because all questions were presented together, participants might have tried to be consistent with their previous answer at the expense of the underlying constructs we aimed to target, possibly resulting in relatively high correlations between all three constructs.

5. Conclusions

According to IRT, a randomized Lifestyle Coach will often recommend interventions above a person's ability level. Therefore, it was expected that an IRT-based Lifestyle Coach should be able to coach more effectively. Two studies have been performed to draw conclusions about the feasibility and effectiveness of IRT-based lifestyle coaching. In the first study we investigated (RQ1) whether it is possible to construct a unidimensional scale of BP-related lifestyle interventions that can be used to assess someone's ability. And, in a second study, we examined (RQ2) whether IRT-based lifestyle coaching can be more effective than coaching without such models.

5.1. Effectiveness of IRT-based Lifestyle Coaching

In Study 1 we demonstrated that it is possible to construct a scale in which BP-related lifestyle interventions are ordered by the challenges they impose, with sufficient unidimensionality for coaching purposes. Study 2 indicates that this scale can be used to coach hypertensive patients more effectively; IRT-based lifestyle coaching through Rasch performed significantly better than coaching at random. As such, the results provide support for the usefulness of IRT models and its application in health recommender systems. The results also suggest it is not required to take into account someone's ability. Tailoring interventions to a person's ability level did not perform significantly better (nor worse) than coaching at random or coaching the simplest interventions that a person did not perform yet. Lifestyle coaching systems should focus on the "low hanging fruits", i.e. the easiest health interventions someone does not apply already.

5.2. Limitations and future work

In the current study the perceived health benefit of a set of lifestyle advices did not seem to be a predictor of the intention to perform these advices. In order to investigate this relationship more thoroughly this study might be extended by not solely presenting recommendations to patients, but also their respective health benefits. Clarifying and informing about the outcome expectations may lead to an enhanced self-efficacy (an important predictor of actual behavior) and enforces a more direct decision making strategy. Furthermore, due to the fact that there was a small negative correlation between the ordering of items along the scale and their objectively measured health benefit, presenting the health benefit of items may result in more preference for the tailored condition. However, an import drawback of labeling behaviors is that people are likely to give a socially desirable response that does not represent their actual intention to perform a behavior.

Although IRT-based lifestyle modelling is a promising technique to aid coaches in developing the right intervention strategy, it should be mentioned that the applied Rasch model poses three important limitations: First, a relatively large sample size is required to make accurate difficulty estimations. Second, because difficulty estimates are based on execution rates it is not suitable for modelling novel, unfamiliar or very specific interventions. And third, the probabilistic model produces a very general scale, in the sense that the item ordering is considered the same for every individual. However, the fact that these behavioral data already fit this general, simplistic and robust Rasch model, and that this model can be used to coach more effectively is very promising.

Future researchers should consider to optimize the personalization of recommendations by using multi-parameter IRT models or separate models for subgroups of individuals (e.g. gender, lifestyle advice, blood pressure, demographics, etc.). Wearable devices, such as smart watches, extract a lot of data from the user which can support this personalization. Also, adding an item discrimination parameter to the item difficulty parameter might already produce a more personalized model. An item discrimination parameter allows researchers to model the degree to which responses to an item vary among ability levels. In other words, this parameter takes into account the item ordering at different ability levels, enabling the construction of a separate scale for each ability level. For instance, for our data this would imply that interventions related to sodium considerations when buying products would be higher on the scale for low ability persons. This requires, however, the use of a 2-parameter IRT model, since the 1-parameter Rasch model is insufficient for these analyses. Important drawbacks of using a 2-parameter model are an increase in complexity and that it requires a large sample size in order to effectively discern between ability levels. This limitation also holds for separate models that might be created based on person characteristics, and may partly explain why splitting the current data by the use of such characteristics did not lead to significantly better results.

In real world coaching systems and applications the IRT model might be dynamically updated by querying users' engagement in lifestyle habits. For instance, if users have to answer one question per week, this allows researchers to fairly quickly build very rich models while the burden for users is small. And, although the amount of participants required for scientific IRT-based experiments is relatively high, this is nothing compared to the amount of users that use m-health applications.

5.3. Practical implications

IRT models can be used to support lifestyle coaches and coaching systems with providing achievable recommendations. It is a promising technique to overcome the cold start problem, a well-known issue in recommender systems, because it provides an estimate of achievable lifestyle advice without the need of personal data. A potential coaching scenario in the context of gradual coaching would be to first, order health behaviors according to the IRT model and to query whether someone already performs an intervention by going up the list from bottom to top. Second, select the easiest intervention someone does not perform, this is probably a good intervention to start with. Third, if a coachee has managed to successfully make a habit out of this behavior, recommend the second best intervention for the upcoming period. This process may continue until all doable interventions from the list have been applied or until a health goal has been reached.

Although the results of the current study suggest that an ability estimate is not required to provide suitable advice, a person's ability might actually be used to support activity classification algorithms. A property of IRT models is that they estimate the likelihood that a person performs a certain behavior based on his or her ability. On top of activity recognition algorithms, the probability of engagement produced by the IRT model might be used as an additional parameter to determine the likelihood that a person performs a certain behavior. If, for instance, a classification algorithm is uncertain about whether someone performs an intervention, that is very difficult for this person according to the IRT model, one may put more emphasis on the likelihood provided by the model.

That lifestyle habits from various health domains (salt-intake, diet and physical activity) form a reasonable uniform construct, suggests it is possible to estimate an ability parameter for various health domains by knowing the performance in only one behavior class. This makes it, for instance, possible to estimate the likelihood of engagement in dietary interventions solely by querying an individual's ability in the stress (Byrka & Kaiser, 2013) or exercise domain. In the context of unobtrusive recordings from wearable devices, such as smart watches, estimating these lifestyle related user characteristics may help to build a more complete user profile. Future research should explore the possibility of estimating a person's ability unobtrusively through recordings of physiological data (e.g. calories burned).

REFERENCES

AHA. (2014). *Understanding and managing blood pressure*. American Heart Association. Retrieved from https://www.heart.org/idc/groups/heart-public/@wcm/@hcm/documents/downloadable/ucm_461840.pdf

AHA [a]. (2015, May). Frequently Asked Questions (FAQs) About Sodium. Retrieved from http://www.heart.org/HEARTORG/GettingHealthy/NutritionCenter/HealthyEating/Frequently-Asked-Questions-FAQs-About-Sodium_UCM_306840_Article.jsp

AHA [b]. (2015, October). Saturated Fats. American Heart Association. Retrieved from http://www.heart.org/HEARTORG/GettingHealthy/NutritionCenter/HealthyEating/Saturated-Fats_UCM_301110_Article.jsp#.Vjjrd5iFMiQ

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., … Emplaincourt, P. O. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, *32*(9; SUPP/1), S498–S504.

Appel, L. J., Brands, M. W., Daniels, S. R., Karanja, N., Elmer, P. J., & Sacks, F. M. (2006). Dietary Approaches to Prevent and Treat Hypertension: A Scientific Statement From the American Heart Association. *Hypertension*, *47*(2), 296–308. http://doi.org/10.1161/01.HYP.0000202568.01167.B6

Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology*, *15*(2), 265–288. http://doi.org/10.1348/135910709X461752

Bandura, A. (2004). Health Promotion by Social Cognitive Means. *Health Education & Behavior*, *31*(2), 143–164. http://doi.org/10.1177/1090198104263660

Beevers, G., Lip, G. Y., & O'Brien, E. (2001). ABC of hypertension: the pathophysiology of hypertension. *BMJ: British Medical Journal*, *322*(7291), 912.

Blair, S. N., Kampert, J. B., Kohl, H. W., Barlow, C. E., Macera, C. A., Paffenbarger, R. S., & Gibbons, L. W. (1996). Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *Jama*, *276*(3), 205–210.

Blokstra, A., Vissink, P., Venmans, L., Holleman, P., van der Schouw, Y. T., Smit, H. A., & Verschuren, W. M. M. (2011). Nederland de Maat Genomen, 2009-2010: Monitoring van risicofactoren in de algemene bevolking. *RIVM Rapport 260152001*. Retrieved from http://rivm.openrepository.com/rivm/handle/10029/256664

Bond, T. G., & Fox, C. M. (2010). *Applying the Rasch model: fundamental measurement in the human sciences* (2. ed., reprinted). New York: Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer Netherlands.

Byrka, K. (2009). *Attitude-behavior consistency: Campbell's paradigm in environmental and health domains* (Dissertation). Eidnhoven, The Netherlands: University of Eindhoven.

Byrka, K., & Kaiser, F. G. (2013). Health performance of individuals within the Campbell paradigm. *International Journal of Psychology*, *48*(5), 986–999. http://doi.org/10.1080/00207594.2012.702215

Chiarini, G., Ray, P., Akter, S., Masella, C., & Ganz, A. (2013). mHealth Technologies for Chronic Diseases and Elders: A Systematic Review. *IEEE Journal on Selected Areas in Communications*, *31*(9), 6–18. http://doi.org/10.1109/JSAC.2013.SUP.0513001

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., … Oparil, S. (2004). *The seventh report of the joint national committee on Prevention, detection, evaluation, and treatment of high blood pressure* (Prevention, Detection, Evaluation, and Treatment of HIgh Blood Pressure No. 7) (p. 86). U.S. department of health and human services. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK9621/

Consumentenbond. (2013). Sandwichspreads, kroketten en kaas steeds zouter. Retrieved from http://www.consumentenbond.nl/actueel/nieuws/nieuwsoverzicht-2013/zoutgehalte-in-voeding-niet-gedaald-sinds-2007/

Dickinson, H. O., Mason, J. M., Nicolson, D. J., Campbell, F., Beyer, F. R., Cook, J. V., … Ford, G. A. (2006). Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. *Journal of Hypertension*, *24*(2), 215–233.

Dima, A., Gibbons, C., Kleppe, M., Byrka, K., Bruin, M., & Johnston, M. (2014). The opportunities Item Response Theory (IRT) offers to health psychologists. *The European Health Psychologist*, *16*(6), 249–259.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists* (1st ed.). New York, NY: Psychology press.

Erhardt, L., Roijer, C., Stagmo, M., & Uden, G. (2004). Better knowledge improves adherence to lifestyle changes and medication in patients with coronary heart disease. *European Journal of Cardiovascular Nursing*, *3*(4), 321–330. http://doi.org/10.1016/j.ejcnurse.2004.05.002

Hallal, P., C., & Lee, I., M. (2013). Prescription of physical activity: An undervalued intervention. *The Lancet*, *381*, 356–357. http://doi.org/10.1016/S0140-6736(12)61426-3

Henson, S., Blandon, J., & Cranfield, J. (2010). Difficulty of healthy eating: A Rasch model approach. *Social Science & Medicine*, *70*(10), 1574–1580. http://doi.org/10.1016/j.socscimed.2010.01.037

Héroux, M., Janssen, I., Lee, D., Sui, X., Hebert, J. R., & Blair, S. N. (2012). Clustering of Unhealthy Behaviors in the Aerobics Center Longitudinal Study. *Prevention Science*, *13*(2), 183–195. http://doi.org/10.1007/s11121-011-0255-0

Hildebrandt, V. M., Bernaards, C. M., & Stubbe, J. H. (2013). Trendrapport bewegen en gezondheid 2010-2011. De Brink, Leiden.

Hill, J. O. (2009). Can a small-changes approach help address the obesity epidemic? A report of the Joint Task Force of the American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *American Journal of Clinical Nutrition*, *89*(2), 477–484. http://doi.org/10.3945/ajcn.2008.26566

Hoeymans, N., van Loon, A. J. M., van den Berg, M., Harbers, M. M., Hilderink, H. B. M., van Oers, J. A. M., & Schoenmaker, C. G. (2014). *Een gezonder Nederland: kernboodschappen van de Volksgezondheid Toekomst Verkenning 2014*. Bilthoven: RIVM. Retrieved from http://www.rivm.nl/dsresource?objectid=rivmp:251654&type=org&disposition=inline&ns_nc=1

Hsee, C. K., & Zhang, J. (2010). General Evaluability Theory. *Perspectives on Psychological Science*, *5*(4), 343–355. http://doi.org/10.1177/1745691610374586

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's Paradigm for Attitude Research. *Personality and Social Psychology Review*, *14*(4), 351–367. http://doi.org/10.1177/1088868310366452

Kaplan, N. M., Victor, R. G., & Flynn, J. T. (2015). *Kaplan's clinical hypertension* (Eleventh edition). Philadelphia: Wolters Kluwer.

Kleppe, M., Lacroix, J., Ham, J., & Midden, C. (2015). The development of the ProMAS: a Probabilistic Medication Adherence Scale. *Patient Preference and Adherence*, 355. http://doi.org/10.2147/PPA.S76749

Lee, M., Saver, J. L., Chang, B., Chang, K.-H., Hao, Q., & Ovbiagele, B. (2011). Presence of baseline prehypertension and risk of incident stroke A meta-analysis. *Neurology*, *77*(14), 1330–1337.

Lichtenstein, A. H. (2006). Diet and Lifestyle Recommendations Revision 2006: A Scientific Statement From the American Heart Association Nutrition Committee. *Circulation*, *114*(1), 82–96. http://doi.org/10.1161/CIRCULATIONAHA.106.176158

Linacre, J. M. (2015). *A user's guide to winsteps ministep Rasch-Model Computer Programs* (3.90.0 ed.).

Lin, J. S., O'Connor, E., Whitlock, E. P., & Beil, T. L. (2010). Behavioral counseling to promote physical activity and a healthful diet to prevent cardiovascular disease in adults: a systematic review for the US Preventive Services Task Force. *Annals of Internal Medicine*, *153*(11), 736–750.

Mancia, G. (2012). Short- and Long-Term Blood Pressure Variability: Present and Future. *Hypertension*, *60*(2), 512–517. http://doi.org/10.1161/HYPERTENSIONAHA.112.194340

Mancia, G., Fagard, R., Narkiewicz, K., Redon, J., Zanchetti, A., Bohm, M., … De Backer, G. (2013). 2013 ESH/ESC Guidelines for the management of arterial hypertension. *Journal of Hypertension*, *31*, 1281–1357.

Meek, G. E., Ozgur, C., & Dunning, K. (2007). Comparison of the t vs. Wilcoxon Signed-Rank Test for Likert Scale Data and Small Samples. *Journal of Modern Applied Statistical Methods*, *6*(1), 10.

Mendoza, J. E., Schram, G. A., Arcand, J., Henson, S., & L'Abbe, M. (2014). Assessment of consumers' level of engagement in following recommendations for lowering sodium intake. *Appetite*, *73*, 51–57. http://doi.org/10.1016/j.appet.2013.10.007

Mika, S. (2011). Challenges for Nutrition Recommender Systems. In *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, Germany* (pp. 25–33). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.415.4084&rep=rep1&type=pdf

Mozaffarian, D., Fahimi, S., Singh, G. M., Micha, R., Khatibzadeh, S., Engell, R. E., … Powles, J. (2014). Global Sodium Consumption and Death from Cardiovascular Causes. *New England Journal of Medicine*, *371*(7), 624–634. http://doi.org/10.1056/NEJMoa1304127

Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M. G., & Poulter, N. R. (2001). Association between smoking and blood pressure evidence from the health survey for England. *Hypertension*, *37*(2), 187–193.

Schippers, E. (2011). *Landelijke nota gezondheidsbeleid*. Retrieved from http://www.regionaalkompas.nl/object_binary/o12028_5258-B5-GezDichbij-binw[web][1].pdf

Sick, J. (2011). Rasch Measurement and Factor Analysis. *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, *15*(1), 15–17.

Sparrenberger, F., Cichelero, F. T., Ascoli, A. M., Fonseca, F. P., Weiss, G., Berwanger, O., … Fuchs, F. D. (2009). Does psychosocial stress cause hypertension? A systematic review of observational studies. *Journal of Human Hypertension*, *23*(1), 12–19.

Starke, A. (2014). *With a little help from my friends: investigating the effectiveness of Rasch-based recommendations with social endorsements* (Thesis). Technical University of Eindhoven, Eindhoven.

Svetkey, L. P., Harsha, D. W., Vollmer, W. M., Stevens, V. J., Obarzanek, E., Elmer, P. J., … Appel, L. J. (2003). Premier: a clinical trial of comprehensive lifestyle modification for blood pressure control: rationale, design and baseline characteristics. *Annals of Epidemiology*, *13*(6), 462–471. http://doi.org/10.1016/S1047-2797(03)00006-1

Uzun, S., Kara, B., Yokusoglu, M., Arslan, F., Yilmaz, M. B., & Karaeren, H. (2009). The assessment of adherence of hypertensive individuals to treatment and lifestyle change recommendations. *Anadolu Kardiyol Derg*, *9*(2), 102–109.

Van Rossum, C. T. M., Buurma-Rethans, E. J. M., Fransen, H. P., Verkaik-Kloosterman, J., & Hendriksen, M. A. H. (2012). Zoutconsumptie van kinderen en volwassenen in Nederland: Resultaten uit de Voedselconsumptiepeiling 2007-2010. *RIVM Rapport 350050007*. Retrieved from http://rivm.openrepository.com/rivm/handle/10029/261552

Van Rossum, C. T. M., Fransen, H. P., Verkaik-Kloosterman, J., Buurma-Rethans, E. J. M., & Ocké, M. C. (2011). *Dutch national food consumption survey* (RIVM-rapport No. 350070001). Bilthoven.

Viera, A. J., Kshirsagar, A. V., & Hinderliter, A. L. (2008). Lifestyle modifications to lower or control high blood pressure: is advice associated with action? The behavioral risk factor surveillance survey. *The Journal of Clinical Hypertension*, *10*(2), 105–111.

Voedingscentrum. (2013). *Het nieuwe eten personal coach: stap voor stap gezonder eten*. Den Haag: Voedingscentrum. Retrieved from https://webshop.voedingscentrum.nl/pdf/050.pdf

Wen, C. P., Wai, J. P. M., Tsai, M. K., Yang, Y. C., Cheng, T. Y. D., Lee, M.-C., … Wu, X. (2011). Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *The Lancet*, *378*(9798), 1244–1253. http://doi.org/10.1016/S0140-6736(11)60749-6

Whelton, S. P., Chin, A., Xin, X., & He, J. (2002). Effect of Aerobic Exercise on Blood Pressure: A Meta-Analysis of Randomized, Controlled Trials. *Annals of Internal Medicine*, *136*(7), 493–503. http://doi.org/10.7326/0003-4819-136-7-200204020-00006

WHO. (2013). *A global brief on hypertension* (No. WHO/DCO/WHD/2013.2). World Health Organization.

Wilson, R., Komitopoulou, E., & Incles, M. (2012). Evaluation of technological approaches to salt reduction. Retrieved from http://appgfooddrink.org.uk/resources/salt_reduction_2012.pdf

APPENDIX A – ITEM STATISTICS

Table 11. Item statistics. Difficulty estimates, fit-statistics, loadings on first PCA contrast and relative health benefit (within the corresponding behavior class).

Lifestyle recommendations (ID's)	Measure	s.e.	Infit	Outfit	Factor loading	Health rank
Dietary						
1	2.56	0.27	1.13	1.28	-0.10	1
2	2.29	0.24	0.97	0.87	-0.17	-
3	2.05	0.22	1.13	1.53	0.23	1
4	1.74	0.20	1.16	1.25	-0.03	1
5	1.33	0.17	1.03	0.91	0.00	1
6	0.82	0.15	0.94	0.86	-0.17	2
7	0.44	0.14	0.97	0.98	-0.05	3
8	0.43	0.14	0.91	0.86	-0.32	-
9	0.03	0.14	1.06	1.08	-0.13	1
10	-0.06	0.13	1.23	1.28	0.15	1
11	-0.15	0.15	0.99	0.98	-0.17	3
12	-0.18	0.13	1.01	1.02	-0.18	2
13	-0.37	0.13	1.26	1.34	0.28	1
14	-0.66	0.15	0.98	0.96	-0.13	3
15	-1.11	0.13	0.90	0.87	-0.18	2
16	-1.19	0.13	0.90	0.87	-0.21	1
17	-1.25	0.13	0.91	0.90	-0.02	2
18	-1.89	0.15	1.14	1.28	0.16	1
19	-2.57	0.18	1.07	1.14	0.18	1
20	-2.65	0.19	1.10	1.24	0.24	1
21	-4.61	0.39	1.01	1.08	-0.01	1
Sodium-intake						
22	2.89	0.32	0.85	0.63	-0.29	1
23	2.74	0.30	0.88	0.57	-0.23	7
24	1.47	0.18	1.00	1.04	-0.22	4
25	1.41	0.18	0.94	0.94	-0.20	5
26	1.10	0.16	0.89	0.71	-0.45	-
27	1.09	0.19	0.87	0.74	-0.23	1
28	0.99	0.16	0.99	1.03	-0.13	2
29	0.81	0.15	0.86	0.72	-0.43	-
30	0.65	0.15	0.94	0.88	-0.21	5
31	0.54	0.14	0.91	0.89	-0.40	1
32	0.28	0.14	0.90	0.84	-0.42	3
33	-0.02	0.13	0.82	0.77	-0.44	1
34	-0.11	0.13	1.01	1.06	-0.29	6
35	-0.37	0.13	0.93	0.93	-0.26	1
36	-0.50	0.15	0.91	0.92	-0.17	7
37	-0.89	0.13	0.91	0.90	-0.30	1
38	-0.95	0.13	0.90	0.87	-0.34	7
39	-1.19	0.16	1.01	1.03	0.02	7
40	-1.55	0.14	0.97	0.94	-0.24	1
41	-2.28	0.20	0.88	0.80	-0.12	1
42	-2.40	0.17	1.13	1.27	0.23	7

Exercise						
43	3.02	0.33	0.91	1.18	0.01	4
44	2.71	0.31	1.04	0.81	0.10	-
45	1.90	0.22	0.96	0.91	0.09	-
46	1.68	0.21	0.97	0.92	0.07	5
47	1.16	0.16	1.05	0.97	0.11	2
48	1.08	0.16	1.01	0.98	0.27	1
49	0.41	0.14	1.07	1.04	0.17	-
50	0.28	0.14	0.92	0.90	0.08	-
51	0.01	0.13	0.98	0.97	0.18	3
52	-0.25	0.13	0.99	0.95	0.24	3
53	-0.28	0.13	1.17	1.23	0.25	4
54	-0.37	0.13	1.07	1.05	0.46	1
55	-0.39	0.13	1.10	1.13	0.14	9
56	-0.44	0.13	1.07	1.11	0.37	1
57	-0.71	0.13	1.10	1.11	0.48	1
58	-0.80	0.13	1.05	1.05	0.24	3
59	-1.09	0.13	1.01	1.01	0.51	7
60	-1.38	0.14	1.06	1.11	0.46	7
61	-1.49	0.14	0.95	0.91	0.30	6
62	-1.53	0.14	0.96	0.96	0.16	8
63	-2.22	0.16	1.07	1.12	0.13	-

APPENDIX B – CALORIC INTAKE

Calorie needs were determined by gender, age and physical activity level. Less than 2.5 hours of physical activity per week was considered *sedentary*, 2.5 to 4 hours *moderately active* and 4 hours *active*.

Table 12. Daily calorie needs [3]

Daily caloric needs for Women

Age (years)	Calories Needed for Sedentary Activity Level	Calories Needed for Moderately Active Activity Level	Calories Needed for Active Activity Level
40-49	1800	2000	2200
50-59	1600	1800	2000–2200

Daily calorie needs for Men

Age (years)	Calories Needed for Sedentary Activity Level	Calories Needed for Moderately Active Activity Level	Calories Needed for Active Activity Level
40-49	2200	2400–2600	2800–3000
50-59	2000	2200–2400	2400–2800

Table 13. DASH serving sizes for three different calorie needs [3]

Food group	**Servings per day***		
	1600 to 1800 Cal.	**2000 to 2200 Cal.**	**2400 to 3000 Cal.**
Grains	4-6	6-9	8-11
Vegetables	4 or more	4 or more	4 or more
Fruits	4-5	4-5	4-7
Dairy products	2-3	2-3	2-3
Meat poultry and fish	3 or less	5 or less	7 or less
Fats and oils	2-3	2-3	2-3
Nuts, seeds and legumes	2-3	4-5	6-9
Sweets and added sugars	3 or less	5 or less	13 or less

* Items in italic are per week

[3] *Retrieved and adapted from: http://www.nhlbi.nih.gov/health/health-topics/topics/dash/followdash#*

APPENDIX C – ITEM MAP

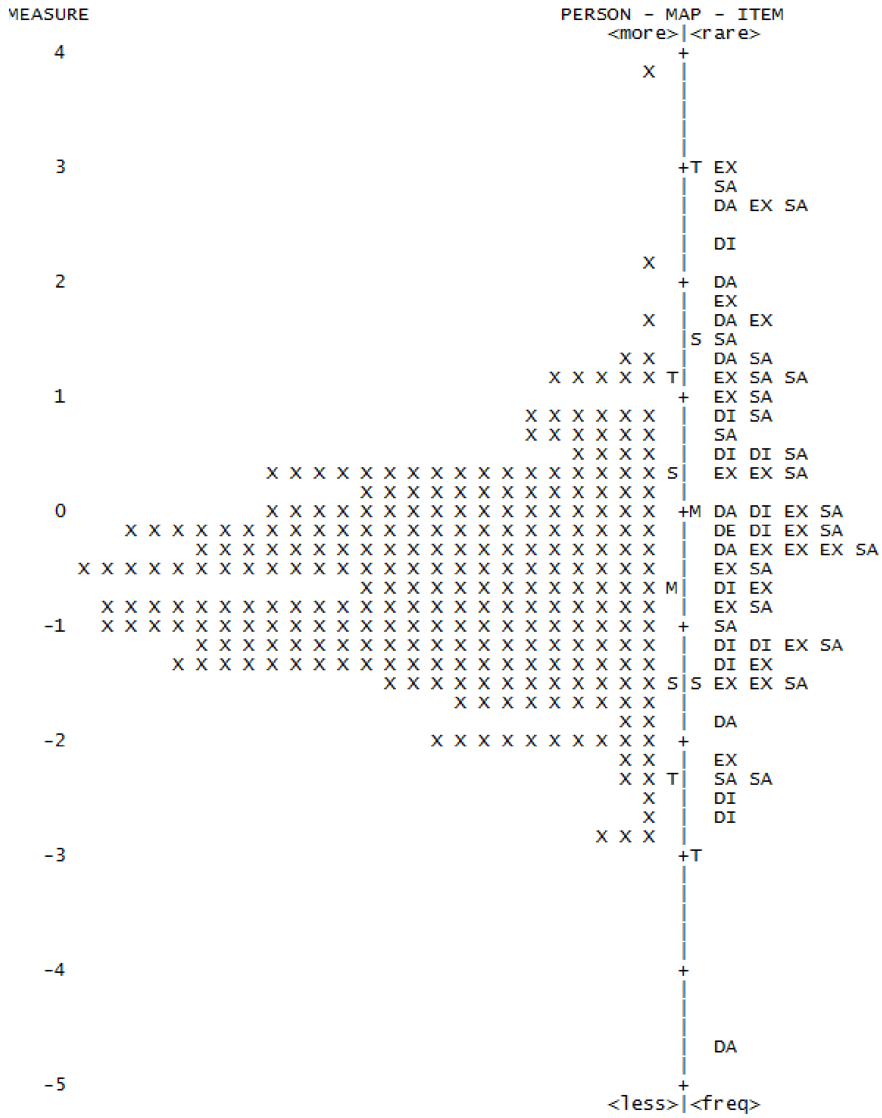


Figure 8. Item map. 285 persons, 63 items. Exercise items (EX), DASH items (DA), Diet items (DI) and sodium items (SA). An X represents 1 person. If a person's ability matches an item's difficulty, there is a 50% chance this person endorses the item.

APPENDIX D – CORRELATION MATRIX

Table 14. Correlation matrix

Variable	Ability	Ability sodium	Ability diet	Ability exercise	Elevated bp	HBP stage	HBP time	Cooking freq.	Smoking	Food restr.	Physical limit.	Multiple chronic cond.	Med.	Age	Gender	BMI	Edu.
Ability																	
Ability sodium	.818**																
Ability diet	.789**	.568**															
Ability exercise	.735**	.348**	.367**														
Elevated bp	.028	.219**	.001	-.153**													
HBP stage	-.095	-.116	-.022	-.078	.[c]												
HBP time	-.181*	-.173*	-.093	-.149	.[c]	-.088											
Cooking frequency	.279**	.232**	.307**	.141*	.090	-.031	.032										
Smoking	-.223**	-.186**	-.207**	-.142*	-.103	.281**	-.087	-.019									
Food restriction	.065	.141*	.057	-.060	.071	-.007	-.129	-.021	.105								
Physical limitations	.019	.125*	.046	-.135*	.196**	.120	.039	.051	.127*	.141*							
Multiple chr. cond.	.060	.138*	.091	-.064	.332**	.049	.082	.044	-.010	.101	.278**						
Medication	.014	.213**	-.020	-.160**	.878**	.021	.398**	.078	-.102	.082	.183**	.313**					
Age	.006	.072	.005	-.070	.276**	.065	.177*	.027	-.023	-.006	.142*	.079	.279**				
Gender	.182**	.193**	.245**	.020	.098	.090	-.151	.316**	-.126*	.150*	.110	.129*	.101	-.047			
BMI	-.104	-.037	-.004	-.197**	.222**	-.018	.216*	-.077	.002	-.010	.080	.236**	.192**	.010	-.067		
Education	.161**	.174**	.135*	.071	.025	-.033	-.117	.094	-.025	-.069	-.017	-.039	-.015	.005	-.059	-.073	
Lifestyle advice	.209**	.343**	.196**	-.060	.414**	-.053	.140	.000	-.114	.266**	.090	.275**	.390**	.098	.057	.196**	.023

* $p < .050$, ** $p < .010$

APPENDIX E – CROSS PLOTS

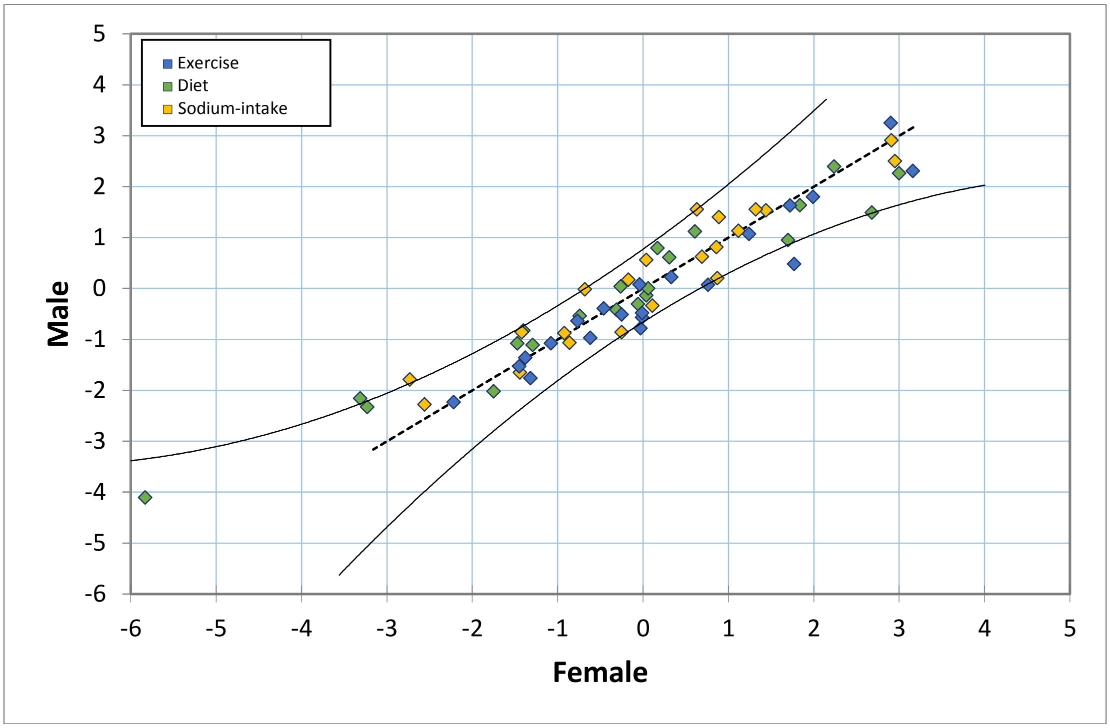


Figure 9. Cross-plot of item measures between males (N=139) and Females (N=146). The diagonal dashed line (slope = 1) represents the line if the item measures were completely invariant. The two solid lines represent 99% confidence intervals.

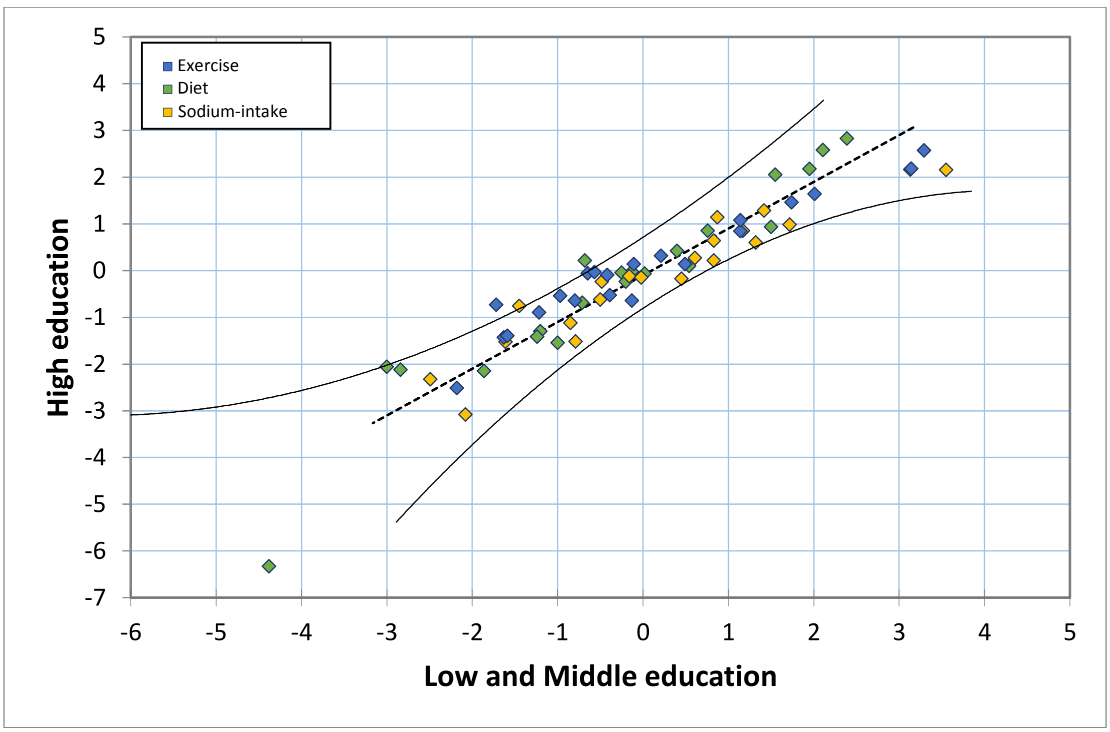


Figure 10. Cross-plot of item measures between high (N=84) and low to medium (N=201) educated people. The diagonal dashed line (slope = 1) represents the line if the item measures were completely invariant. The two solid lines represent 99% confidence intervals.

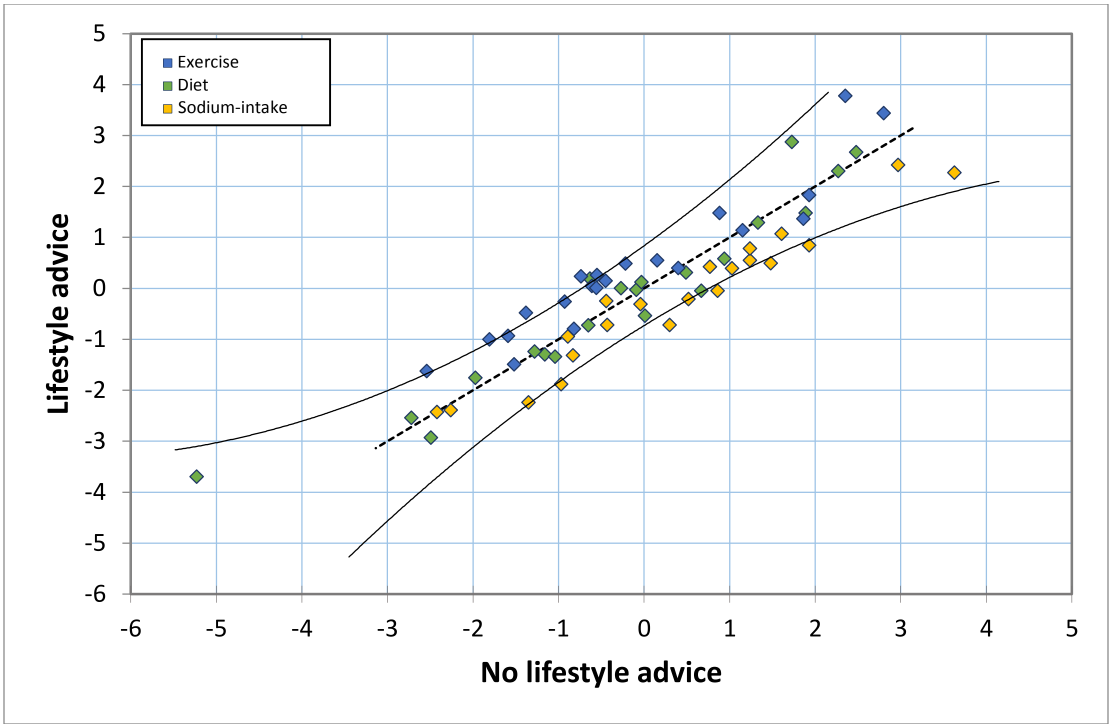


Figure 11. Cross-plot of item measures between people that did (N=88), and did not (N=197), receive lifestyle advice from a healthcare professional. The diagonal dashed line (slope = 1) represents the line if the item measures were completely invariant. The two solid lines represent 99% confidence intervals.

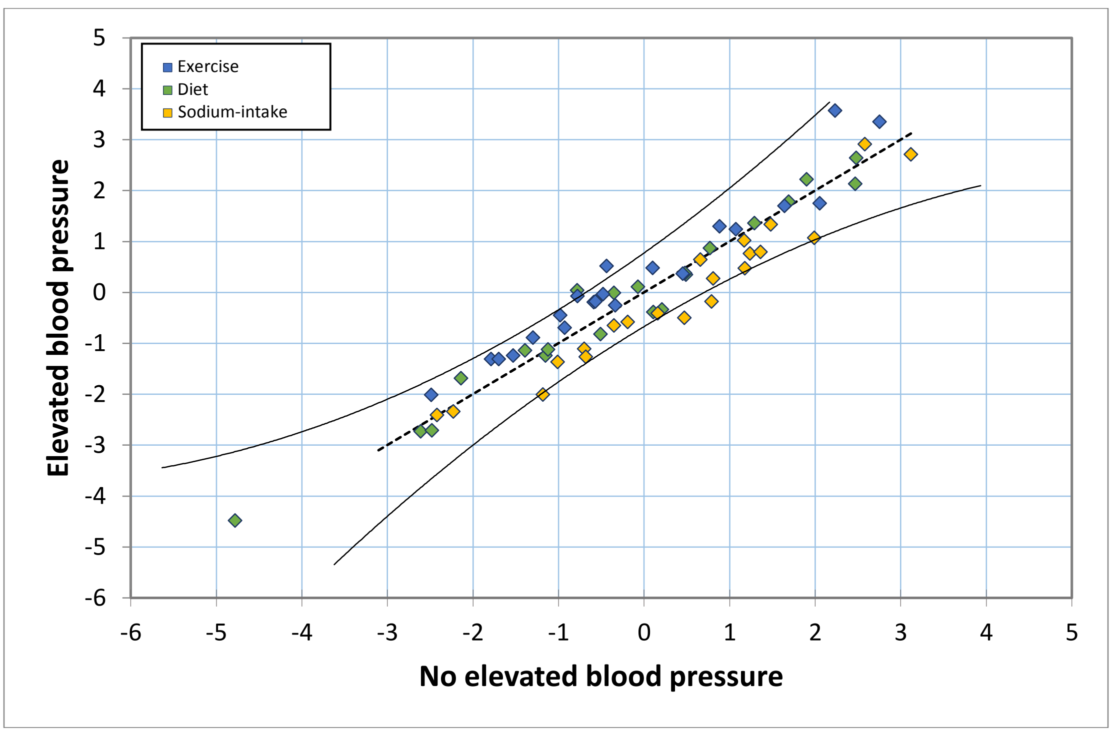
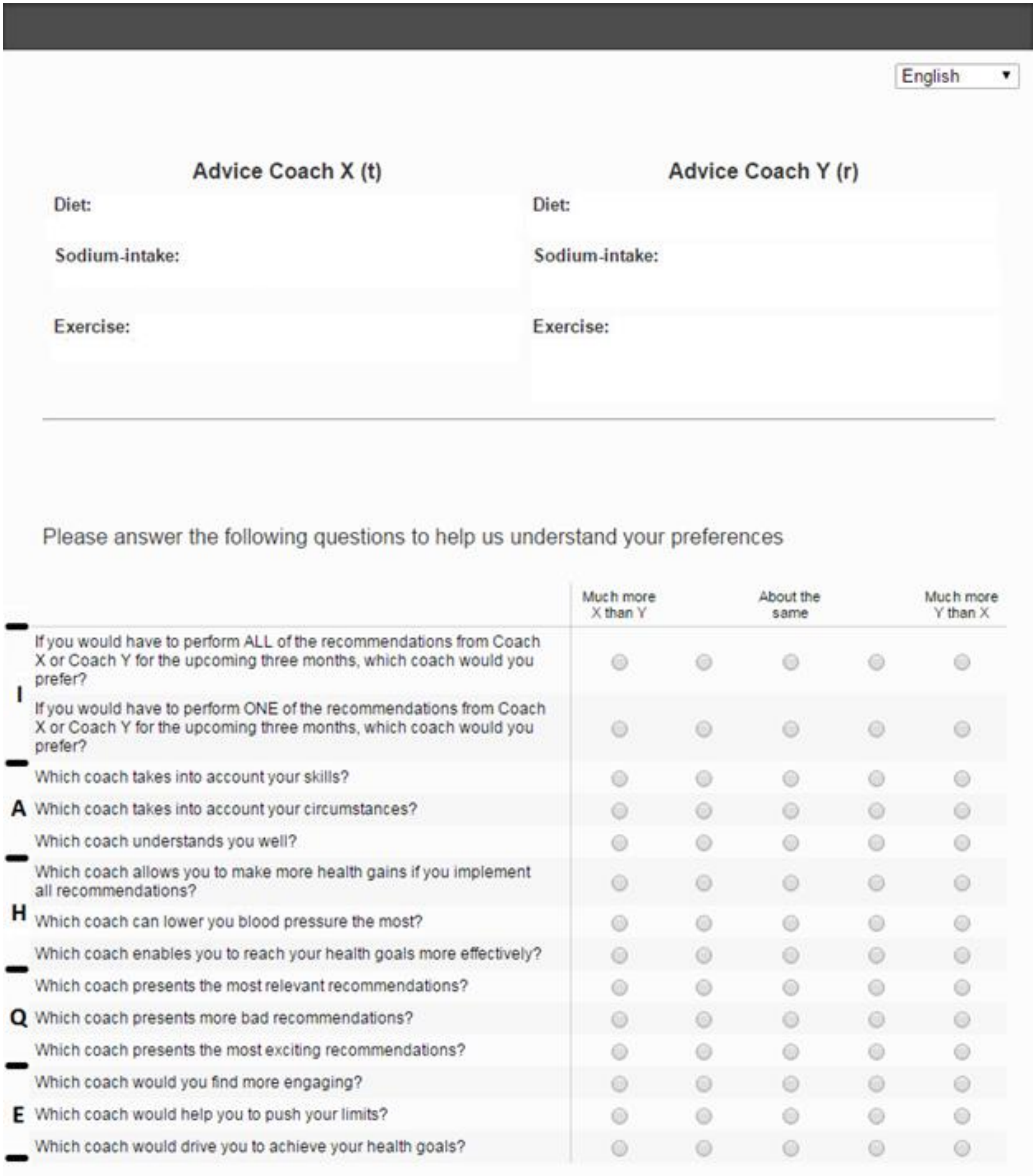


Figure 12. Cross-plot of item measures between people that did (N=142) and did not (N=143) have, or were unaware of having, an elevated blood pressure. The diagonal dashed line (slope = 1) represents the line if the item measures were completely invariant. The two solid lines represent 99% confidence intervals.

APPENDIX F – EXPERIMENTAL SET-UP STUDY 2



English

Advice Coach X (t)

Diet:

Sodium-intake:

Exercise:

Advice Coach Y (r)

Diet:

Sodium-intake:

Exercise:

Please answer the following questions to help us understand your preferences

		Much more X than Y		About the same		Much more Y than X
I	If you would have to perform ALL of the recommendations from Coach X or Coach Y for the upcoming three months, which coach would you prefer?					
	If you would have to perform ONE of the recommendations from Coach X or Coach Y for the upcoming three months, which coach would you prefer?					
A	Which coach takes into account your skills?					
	Which coach takes into account your circumstances?					
	Which coach understands you well?					
H	Which coach allows you to make more health gains if you implement all recommendations?					
	Which coach can lower you blood pressure the most?					
	Which coach enables you to reach your health goals more effectively?					
Q	Which coach presents the most relevant recommendations?					
	Which coach presents more bad recommendations?					
	Which coach presents the most exciting recommendations?					
E	Which coach would you find more engaging?					
	Which coach would help you to push your limits?					
	Which coach would drive you to achieve your health goals?					

Figure 13. Experimental set-up Study 2. Participants had to indicate their preference for one of the two Lifestyle Coaches (all three conditions, i.e. Random, Tailored and Simple, were compared in a similar fashion). The questions were divided into five categories; I = Intention, A = Achievability, H = Health Benefits, Q = Recommendation Quality, E = Engagement.