Indivi

Designing a tool for researchers to create individualized feedback

Master's Thesis

People and Computing Lab
Department of Informatics
University of Zurich

by
Florian Fischer

Supervised by
Prof. Dr. Chat Wacharamanotham
Dr. Andrea B. Horn, Dipl.Psych.

Submission: 31 December 2018

University of Zurich[UZH]

PAC

Contents

List of Figures

Abstract

English

In the field of Psychology, researchers typically conduct studies with human participants. With the raise of new technologies (e.g., smartphones, sensors) and the way they affect humans' everyday life, participants want to be more involved in research, help to shape research questions and receive feedback. To fulfill this expectation, researchers in Psychology are in need of a tool that guides them and automates the creation of feedback reports for studies. This research focuses on three research questions: (1) find out the design principles for such a tool to follow the current practice of giving feedback, (2) find out the design principles for such a tool to output individualized feedback, and (3) infer a good visualization for a specific feedback given the type of data provided. This research project elicited the requirements and developed a tool called *Indivi*. (1) *Indivi* creates individualized feedback following a workflow that was elaborated together with researchers in Psychology. (2) *Indivi* individualizes feedback on three levels: individualizing graphs displaying the individual value, individualizing texts by using placeholders to input the individual value, and individualizing text by classifying participants into categories and writing a text for each category. (3) *Indivi* also infers a fitting type of graph based on the type of data provided and some user choices. *Indivi* was found to represent an improvement compared to manual practices. In future developments, *Indivi* would benefit from being extended to a platform not only for researchers but also for participants. It could provide instant feedback generated directly from the data collection and this feedback be used to steer behavioral changes.

Deutsch

In der Psychologieforschung liegt der Fokus liegt typischerweise auf Menschen als Probanden. Neue Technologien wie Smartphones oder Sensoren verändern den menschlichen Alltag und die Erwartungen von Probanden an die Forschung hat sich weiterentwickelt. Unter anderem erwarten Probanden Feedback für ihre Teilnahme. Forschende benötigen dafür

ein Tool, welches die Entstehung dieses Prozesses leitet und automatisiert. Dieses Projekt basiert auf drei Forschungsfragen: (1) Wie wird heute Feedback an Probanden generiert und welchen Gestaltungsprinzipien muss ein solches Tool unterliegen? (2) Welchen Gestaltungsprinzipien unterliegen die Individualisierung von Studienfeedback? und (3) Wie kann eine passende Grafik von den eingegebenen Daten abgeleitet werden? Dieses Forschungsprojekt hat die Nutzerbedürfnisse eruiert und ein Tool namens *Indivi* entwickelt. (1) *Indivi* generiert individualisiertes Feedback für Probanden gemäss einem Prozess, welches gemeinsam mit Forschenden der Psychologie entstand. (2) *Indivi* individualisiert Feedback auf drei Ebenen: individualisierte Grafiken, die das eigene Testresultat der Probanden aufzeigen, individualisierte Texte, in welchen Platzhalter für die Beschreibung der eigenen Testresultate eingesetzt werden, sowie die Klassifizierung von Probanden in Kategorien, für welche passende Texte geschrieben werden. (3) *Invidi* leitet auch die passende Grafik von den eingegebenen Daten ab. Forschende der Psychologie, die *Indivi* testeten, fanden es eine Verbesserung im Vergleich zu manuellen Anwendungen. Weiterführende Arbeit könnte *Indivi* zu einer Plattform erweitern, die ebenfalls von Probanden selber genutzt werden könnte. Feedback könnte dadurch direkt während der Datenerhebung generiert und unmittelbar an Probanden zugestellt werden, beispielsweise um Verhaltensveränderungen zu untersuchen.

Acknowledgements

This research would not have been possible without the implication and the engagement of various actors who provided continuous support and insights over the six months of this research project.

Thank you to Prof. Chat Wacharamanotham who provided guidance and feedback throughout the project, especially on the methodology of the need-finding phase, on the user-centricity and user experience of the implementation and on the writing-up of the research results.

Thank you to Dr. Andrea Horn who initiated this project with Prof. Wacharamanotham and supervised the research from the user side. She played the role of the principal and participated in all the steps of the process as user. She also opened her network to recruit users and testers outside her lab and even outside the University of Zurich.

Thank you to Marko Katana who participated with a lot of engagement in every step of the process, providing feedback, developing ideas and sharing his experience.

Thank you to Prof. Emily Butler, Dr. Corina Berli, Dr. Susan Merillat, Jasmine Willi, Laura Mernone, Tabea Meier and Walter Bierbauer who participated in the interview and the testings and shared their experience and expectations.

Chapter 1

Introduction

The science of Psychology has historically been based on experimental and observational studies, similarly to other scientific fields such as Physics, Chemistry or Biology. As research in this field is interested in psychological processes and behaviors, studies are conducted with human beings and therefore underlay strict ethical rules. The essence of the work of researchers in Psychology is to design, conduct and analyze studies [Fiske et al., 2010].

Expectation of receiving feedback

With the increasing use of smartphones and the availability of sensors, data can be collected continuously and feedback has become a natural part of daily life in a large number of context such as health apps, hotel ratings, restaurant reviews, school test, exams or assessments, and many more [van Berkel et al., 2017]. In most of these cases, the feedback received in everyday life is individualized in a way that each participant receives exactly the relevant information. Therefore, researchers are increasingly expected to also provide feedback to the participants of their studies.

Number of participants in studies

As the technologies evolve, the marginal cost of participation of a further participant decreased and researchers tend to include increasingly more participants to their study. Researchers in Psychology face the double challenge of providing feedback and to individualizing the feedback. Researchers are therefore in need of a tool that guide them and automate the process of creating individualized feedback for their research participants.

1.1 Types of studies in Psychology

Temporal aspect of Psychological studies

To understand the challenges of giving feedback about Psychological studies, it is important to understand how studies are constructed and what data is collected. An important characteristic to differentiate between types of studies is their temporal aspect, meaning how many times a participant will be assessed and over which period of time [Ram and Gerstorf, 2009, Bolger et al., 2003]. Four types are mainly used today in the field of Psychology: cross-sectional studies, longitudinal studies, micro-longitudinal studies and studies using the measurement burst approach. In many areas of Psychology, most studies are cross-sectional [Ram and Gerstorf, 2009].

Cross-sectional studies

Cross-sectional studies have been the most used format of studies, being simple and cheap as they include no time component. Participants are required to take part in the study once and all assessments included in the study are conducted at the same time point. Cross-sectional studies show the performance of participants on a particular day and time. The results of such studies represent a snapshot of a momentary trait of the participants, independently of their current abilities, emotions or other external factors [Bolger et al., 2003, Shiffman et al., 2008]. These studies collect only one value for each assessment and each participant. Participants can therefore not be compared with themselves (within-person) but only with other participants (between-person). An example for a cross-sectional question could be: "Please rate on a scale from 1 to 10 how happy you were during the last month".

Longitudinal studies

Longitudinal studies assess participants multiple times at different time points. The assessments are usually the same for each time point and yield multiple results for the same assessment for each participant. Thus, they aim at studying changes over periods of time. Thanks to the availability of multiple data point for each variable and for each participant, the analyses of longitudinal studies can be conducted both on the within-person level, comparing the same individual over time, and the between-person level, comparing individuals at the same time point as well as comparing the individual changes [Curran and Bauer, 2011, Fleeson, 2007, Kievit et al., 2013]. The duration between time points depends on the research question but are usually long, ranging from weeks to years. Between two time points, participants are usually confronted to specific conditions such as training, drugs or simply the effect of time. Although two time points are

already sufficient for a study to qualify as longitudinal, as in studies comparing the state of participants before and after being exposed to a condition, many longitudinal studies are conducted over more time points.

Micro-longitudinal studies

Micro-longitudinal studies are a sub-category of longitudinal studies. They also refer to studies repeating their assessment multiple times with the same participants over time. The main difference lays in the duration between to time points. While, it is quite long in longitudinal studies, the assessments in micro-longitudinal studies repeat every day or even multiple times per day [Ram and Gerstorf, 2009, Nezlek, 2012, Bolger et al., 2003]. This type of study emerged thanks to the wide spread use of smartphone that enable a new type of measurement called ambulatory assessment [Trull and Ebner-Priemer, 2013, Shiffman et al., 2008]. Since it is cumbersome for participants to come by the lab every day or even multiple times per day, the assessments conducted in micro-longitudinal studies are more easily inserted in the real-life of the participants and conducted in a mobile and ambulatory way. Their setting imply repeated measurements per variable and per participant. As in longitudinal studies, micro-longitudinal studies enable to analyze results both on the within-person level and on the between-person level but with a focus on short-term fluctuation rather than long-term change [Curran and Bauer, 2011, Ram and Gerstorf, 2009, Nezlek, 2012]. Micro-longitudinal studies, empowered by new technologies, allow to investigate research questions that were not easy to answer before because the study design were hardly feasible [Ram and Gerstorf, 2009].

Measurement burst studies

As both longitudinal and micro-longitudinal studies have become popular in some areas of Psychology, a combination of both study type has emerged. The so-called "measurement burst" studies repeat over longer periods of time shorter periods during which participants are assessed a large number of times. The measurement burst approach enables to study both long-term changes and short-term fluctuations as well as changes in the individual short-term fluctuation over time or after the exposition to a condition [Sliwinski, 2008]. Figure 1.1 shows how the four different types of study differ in their temporal component with cross-sectional studies taking place only at a single time point, longitudinal studies at multiple time points over longer periods of time, micro-longitudinal studies repeating multiple times over a short period of time and measurement burst combining both [Sliwinski, 2008].

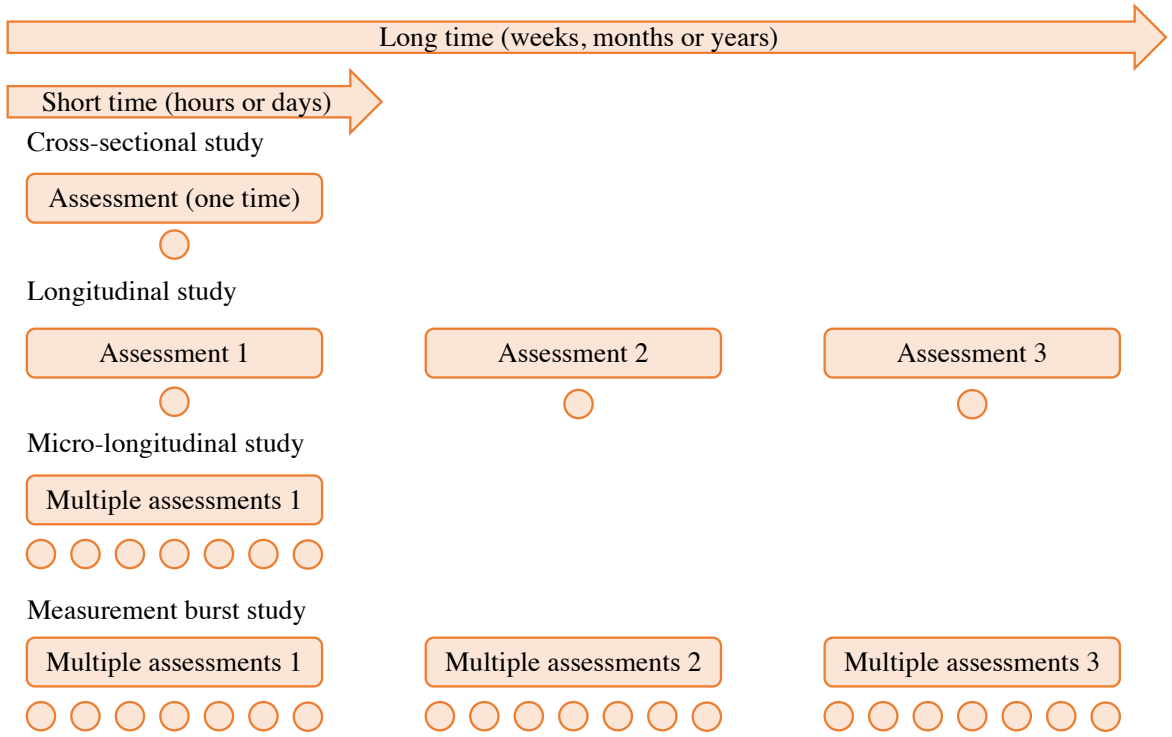


Figure 1.1: Temporal aspect of psychological studies: differences between cross-sectional, longitudinal, micro-longitudinal and measurement burst studies.

Dyads in Psychology

Another characteristic to differentiate between types of studies is whether participants are studied as individuals or as dyads (i.e. pairs). Typical dyads are couples but they can also be of other types such as patient-doctor, parent-child or friendships [Atkins, 2005]. The dyadic component enlarges the possible analyses by explaining results from one individual by the results of his or her dyad. In presenting dyadic results, the values of both dyads must be clear and the relation made visible. The research questions in studies focused on assessing participants in a dyadic setting are somehow different in than studies researching only individuals.

1.2 Types of data in psychological studies

Studies in the field of Psychology collect and analyze various types of data. Although there is no established way of categorizing data in

Psychology, a classification is proposed here to shed light on the challenges of giving feedback in psychological studies. The goal of this section is not to be exhaustive but to give an overview of the types of data gathered in psychological studies for readers with no background in Psychology.

Self-report as assessment measure

When thinking about studies in Psychology, one often pictures someone filling out a questionnaire. This is explained by the fact that numerous studies and areas of Psychology are interested in the self-report of participants [Fiske et al., 2010]. Self-report means the assessment by the participant of himself or herself. Typical items collected through self-report are emotion regulation and personality trait. Self-report is conducted by the use of questionnaires where the researcher has formulated a question based on a reliable and validated scale for the participant to choose from. Such questionnaires can be proposed to the participant either in the presence of the researcher, usually in the lab, or outside the lab during the participant's real life. The questionnaires can be filled on a computer or on paper [Green et al., 2006]. The questionnaires typically assess multiple constructs (e.g. affect or emotion) based on multiple items (i.e. questions). Constructs of interest can be either built from one or from multiple items, in which case the average of the answers is computed. For each construct, a value is then available and ranges within the scale (e.g. from 1 to 7). The values are either discrete if the construct is composed of only one item or continuous if an average was calculated. If the study format includes multiple time points, a value is available for each construct and for each point in time.

Cognitive abilities and capacities

Measuring mental capacities and abilities is referred to as cognition and is also an important task in psychological research. Cognition is measured with the help of psychometric tests that have been specially created and validated by independent studies. Cognition tests can take various forms such as memorizing words and numbers, picture vocabulary test, pattern recognition or list sorting test [Weintraub et al., 2013]. As for self-reports and sensors, the data collected is not automatically the value of interest about which feedback should be given. However, each test is constructed to have a meaning on its own and the researcher has to decide whether to give feedback on the test or to build a value that is more representative of the question of interest. Some cognitive tests are long and require a researcher to conduct the test, making them expensive. Others are quick and simple and can be done on its own. The format of the study therefore puts restrictions on the tests than can be conducted.

Objective measures of behavior

Research in Psychology is also interested in assessing behavior and mental capacities in a more objective way. For measuring the behavior of participants, new technologies have provided new possibilities. In order to understand a participant's mobility, GPS-tracking devices can be used and provide data free from the subjective perception of the participant. Thus, sensors have been increasingly used in recent psychological studies. Other example of sensors include accelerometers, which measure acceleration and give insights about the physical activity, heart rate sensors as well as sweat sensors that can be used to assess stress levels of participants, or audio spinets to evaluate social interactions. There are several difficulties with the sensor data. One difficulty is that the sensors do not directly measure the constructs that the researchers are interested in [Harari et al., 2017]. Scales have to be built from the raw data witch is not as easy as computing an average or sum score of self-report items. Another difficulty is that the values generated by the sensor data do not guarantee to have an interpretable meaning. They are also generated continuously and independently of the measurement time points of the study designed by the researchers. As a result, an important processing work and data cleaning has to be done to enable having one value for each participant, for each item and for each time point [Mehl et al., 2001].

Psychological studies can also include biological values such as neurological tests and images, hormones or genes. These values are not specific to the field of Psychology, have not been explored in the present research, and are therefore not further illustrated. Sensor data is still uncommon in the field of Psychology and more used in other fields of studies. Self-report data and cognition data are the types of data most specific to the field of Psychology. With this understanding of field of Psychology, its types of studies and the data that results from them, it is possible to deep-dive into the subject of the present research - providing feedback to participants - and to motivate the need for such research.

1.3 Motivation

Participants' involvement in Psychological studies

In the activity of researchers in Psychology, recruiting participants play an important role as the studies could not take place without them. With the raise of new technologies, studies can take place over a long time period and require participants to answer questions very frequently while sensor data is being recorded continuously. The

amount of effort required from participants as well as their sharing of very personal data is increasing while the return for participating in the study has not evolved. The present research is motivated by the gap between what researchers ask from participants and what they give in return. This gap is tangible in many ways, including the difficulty to recruit participants for demanding studies and to difficulty to retain these participants over time when they have realized the effort associated with their participation.

Researchers' workload associated with giving feedback

The present research is also motivated by the large workload associated with giving feedback for researchers in Psychology. The workload is kept high because of three factors. First, as there is neither a standard practice for delivering feedback nor clear guidelines on how to prepare the feedback, researchers often have to create their feedback from the scratch, going through difficulties and errors that have often been experienced by other researchers before them. Second, researchers in Psychology often have no training in data visualization and this field can be very challenging when not approached properly. Much energy and efforts are lost trying out visualization formats while best practices are available. Third, the number of participants in Psychological studies tend to increase as the data collection can be automated. If the feedback process is not automated, proving hand-made feedback to hundreds or thousands of participants is highly time consuming and the individualization becomes almost impossible.

Outlook on the up-coming chapters

The present research focuses on eliciting the design requirements and developing a tool allowing researchers in Psychology to create individualized feedback. This research addresses the topics of giving feedback in general, of providing individualized feedback, and of visualizing feedback. This research builds on previous studies and ethical considerations of giving feedback in Psychology as well as on common challenges in visualizing data (chapter 2). In chapter 3, the method of contextual inquiry used in the need-finding process is described and the results of this process are presented with regard to the general practice of providing feedback, to ways of individualizing feedback, and to visualizing data. The elicited requirements have been implemented in a tool for which the design decisions taken are reported in chapter 4. In particular, the flow of the tool, the individualizing features and the chosen visualizations are presented. The technical implementation of the tool is showed in chapter 5. The results of this research as well as the testing feedback are discussed in chapter 6. A conclusion and an outlook is proposed in chapter 7.

Chapter 2

Related work

In this chapter, the results of a literature review in the area of this research are presented. Since this research is about giving individualized and visualized feedback in the field of Psychology, the three topics are treated separately. First, the literature on the practice of giving feedback in Psychology is assessed on a general level and linked to the ethical considerations that underlay the work of any psychologist. Second, the literature was searched with regard to results on individualized feedback. Third, challenges in visualizations were collected and recommendations gathered. From the insights gained from the literature, three research questions are derived and presented. Answers to these research questions provided by this research will be presented in the following chapters 3 and 4. The results will then be discussed and evaluated in the chapter 6.

2.1 Feedback in Psychology: ethical considerations

Participants' expectation and researchers' practice

Every client in a clinical relationship or every potential participant in Psychological study might have been confronted with personalized feedback in various forms. Smartphone utilization statistics, number of steps during a day, carbon footprint for the last flight, feedback is becoming omnipresent and normal thing to expect. However, psychologist, especially in academic settings, are not used to give feedback to their clients, patients or participants and literature on feed-

back practices in Psychology is scarce [Smith et al., 2007, Curry and Hanson, 2010]. 25 years ago, Pope [1992] wrote an article urging researchers in Psychology to conceive feedback as an inherent part of assessment. Since then, only a few results have been published with regard to the best practices, the frequency or the benefits of giving feedback. Because of the lack of publication in the academic (non-therapeutic) context, the focus was broadened to feedback in general, for example in a clinical or therapeutic context [Smith et al., 2007, Shalowitz and Miller, 2008, Curry and Hanson, 2010, Dixon-Woods et al., 2011]. However, Lefaivre and colleagues stress that psychological testing should be differentiated from diagnostic assessments [Lefaivre et al., 2007].

Ethical recommendation from the American Psychological Association

Providing psychological assessment feedback is an ethical obligation. The American Psychological Association (APA) is a world-wide respected body in both clinical and academic psychological practices. One of its area of expertise is in developing ethical principles for psychologists and other scientists and in defining a code of conduct, which is published and publicly available [American Psychological Association, 2017]. Even though the present research focuses on giving feedback in an academic context, it is important to understand the ethical principles that apply to any activity within the field of Psychology and why they are particularly important for feedback in an academic setting. The principle of avoiding harm obliges psychologists to take reasonable steps to not only avoid but also minimize potential harm of any kind. In the research context, this implies ensuring that no harm is induced to the participants before, during but also after the study, making it an ethical obligation to organize, plan and execute feedback as thoroughly as the rest of their research [American Psychological Association, 2017]. Informed consent is another crucial ethical principle in Psychology. It is the obligation for psychologist to receive a consent from their participants. For this consent to be informed, it must happen after the participant has been given all the relevant information presented in a way that let no doubt about its ability to understand. In the same logic, participants have not only the right but the need to be informed, to understand the context and the reason not only before the study but also at any point during the study and after the study. [Pope, 1992, American Psychological Association, 2017]. The most detailed guidance provided by APA is on debriefing, where it is stated that psychologists are ethically obliged to inform the participants about the results and conclusions of their research. Both the way of compensation (including feedback) and informed consent must be reported in most academic psychological journals when pub-

lishing results from a study.

DEBRIEFING :
Psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware ([American Psychological Association, 2017], 8.08).

Ethical recommendation from the Swiss Society of Psychologists

More locally, many other professional societies provide ethical guidelines and codes of conduct. The Swiss Society of Psychologists (SSP), for example also issues a document on ethical principles that apply to its members. On the principle of avoiding harm, SSP adds the ethical obligation of making the necessary efforts to guarantee the well-being of the persons involved in their work. SSP also highlights the right of self determination (Selbstbestimmungsrecht) as core principle. Co-determination is naturally linked to information, leading to informed consent and right to feedback. SSP follows APA in its principles for debriefing, mandating the psychologists to offer the opportunity as soon as possible to get feedback in form of information of the content, the results and the consequences of the research. No indication is made on individualized feedback [Schweizerischen Gesellschaft für Psychologie, 2018].

Benefits for participants

Above and beyond the codes of conduct, respect for others in general and for study participants in this case is per se a sufficient reason to provide feedback and research results [Fernandez et al., 2003]. Besides the ethical considerations and the responsibility that arise thereof, giving feedback is a good practice because it is beneficial for the recipients. It helps them making informed decisions about their lives, objectify their current states and offers a communication base for their families and relatives. It fosters the quality of life and might lead to needed interventions [Smith et al., 2007, Fernandez et al., 2003]. In their study, Smith and colleagues [2007] asked over 700 psychologists (mainly neuropsychologists and clinical psychologists) about the estimated effects of giving feedback. Their results showed that the feedback fostered dialogue, deepened understanding and increased motivation.

Benefits for researchers

Curry and Hanson [2010] showed that beside the benefits for the recipient of the feedback, such as gaining context for understanding or grasping a direction for change, feedback giver also gain valuable insights and understanding on the process of taking the test or participating in the study. Fernandez and colleagues [2003], arguing from

the general perspective of human research, understand feedback as a way for researchers to disseminate their research results and more generally to raise interest and awareness for their research field. For them, feedback shows the central nature of their participation to research, prevent participants to feel exploited and leads to more trust and more engagement of participants in research. They thus suggest considering study participants as a fourth target group of research communication beside the scientific community via conferences and peer-reviewed journals, the general population via lay media and special interest groups via their community or organizations (e.g., elderly organizations, Alzheimer support group, etc). Shalowizt and Miller [2008] highlight that feedback is associated with gratefulness among participants and that this could improve the public perception of research and researchers, bind and retain participants to the study or improve trust. However, they raise concerns that the lack of data on feedback practices and on participants' perception of feedback might infer wrong interpretations on the benefits of feedback [Shalowitz and Miller, 2008].

Risks of feedback

Giving feedback is associated with risk of misunderstanding and misinterpretation [Dixon-Woods et al., 2011]. The way results are presented and the way the feedback is given has a direct influence on how the receiver of the feedback will interpret its meaning and decide whether to take action. So as feedback can go awry, there is a need for proper clarification of expectations, roles and tasks on both the feedback provider side and the feedback recipient side [Pope, 1992]. In analyzing the feedback process of the ORACLE Children's study (OCS), a large longitudinal study in the years 2000 that researched the long term effects of drugs given to their parents, Dixon-Woods and colleagues [2011] found out that the risks associated with giving feedback can be mitigated when properly and holistically designing the feedback process. They recommend to design the feedback process as early as possible, possibly as an inherent part of designing the study. In developing the feedback, a clear focus must be laid on the perspective of the participants, their concerns, expectations and interests. Having a separate team to develop the feedback can be useful because in-depth knowledge of the topic might make taking the participants' perspective more difficult. Consulting with (potential) participants and more generally with lay-persons helps focusing on the participants and finding the right balance between scientific reporting and intelligibility. For large and sensitive feedback, the use of an external website and a telephone hotline might be needed. Since receiving feedback is associated with risks, participants must have a right

to decline feedback [Fernandez et al., 2003, Dixon-Woods et al., 2011, Shalowitz and Miller, 2008, Lefaivre et al., 2007].

Types and forms of feedback

Research that focused on feedback practices in Psychology distinguish between the content of feedback, the type of feedback and the delivery of feedback. The content of feedback can be general or individualized, the type of feedback can be verbal (in person) or written and the delivery of feedback can be direct (to the recipient himself) or indirect to family members or other health professionals (e.g., primary care physician) [Smith et al., 2007]. The most common source of feedback for test assessments was found to be by far verbal feedback, while the importance of having a written trace was stressed [Curry and Hanson, 2010]. Independently of the format of feedback, it must be communicated in a way that is understandable by participants. Considering the distribution of technical reports or of any type for which the recipient has no background to easily grasp the content is considered a bad practice [Smith et al., 2007].

2.2 Giving individualized feedback

Historically, also for technological reasons, individualized feedback was given in person, for example by a clinician to his patient after having conducted tests or assessments. Today, various methods exist to overcome the previous barriers and individualized feedback is becoming more common. In this section, results from the literature are presented with regards first to the ethics and benefits linked to giving individualized feedback and second to the recommendations for providing individualized feedback.

Ethical consideration about individualized feedback

With the raise of data collection enabled by the use of electronic devices in many areas of our everyday life, data privacy and access to data is raising questions and is subject to many ethical considerations. Even though the Ethical Principles and Code of Conduct of the American Psychological Association gives no guidance on the ethical aspects of individualized feedback, it specifies ethically correct behavior in the case of automated feedback. In their interpretation of results, psychologists have the ethical obligation to reflect on the purpose of the assessment or of the research and on all possible factors (situational, personal, linguistic, cultural, etc.) that might have affected the results of the study [American Psychological Association, 2017, Lefaivre et al., 2007]. Between the ethically grounded expectations of participants

and the absence of formal guideline for individualized feedback, a gap exists and needs to be filled [Lefaivre et al., 2007].

Benefits of individualized feedback

While feedback is an ethical obligation, it also carries benefits for the recipients whose expectations towards feedback have increased from simply knowing about general results, as the codes of conduct recommend [American Psychological Association, 2017, Schweizerischen Gesellschaft für Psychologie, 2018], towards debriefing their own assessment and getting access to their own data. In general, receiving feedback is for the participants an opportunity to gain knowledge, which can be understood as counterpart for the participation [Lefaivre et al., 2007]. As it is common in Psychology to tell participants a cover story before a study in order to test placebo effects or usefulness of an intervention, it is especially important to give feedback in studies with covers stories. A review of studies reported that providing personalized information has a positive effect on the recipient, increases his satisfaction and that such effects were less strong if the feedback was general and not individualized [Smith et al., 2007]. Shalowitz and Miller [2008] conducted a qualitative review of studies and articles on the topic of providing feedback (especially in the medical field and not in Psychology) and found that what participants expect from feedback is primarily to see their individual study results and the aggregates of the results of all participants.

Recommendations for individualized feedback

From the literature, some recommendations and best practices emerge. In line with Lefaivre et al. [2007], recommendations are presented with regard to the type of feedback delivery, the time point of feedback delivery, the appropriateness and good practices for delivering undesirable feedback. Then, possible solutions to problems related to individualized feedback are described such as the need for normative and comparative data or its potential coercive effect on participants.

Delivering individualized feedback

Delivering feedback is an important topic that influences the feasibility of providing individualized feedback. In the literature, as explained in the previous section, a difference is made between providing feedback verbally and in person compared to providing a written feedback. The former approach is linked to large costs, especially in the case of large studies, knowing that the in person feedback has to be given by the researchers themselves or by trained experts and is often assumed to be the better option than written feedback[Shalowitz and Miller, 2008, Lefaivre et al., 2007]. When given in person, feedback is a dynamic and interactive process, in which the feedback giver can ad-

just and clarify misunderstandings right away [Pope, 1992]. However, research on participants' feedback has shown that participants would often prefer to receive their feedback in written form jointly with contact information and/or a contact opportunity [Shalowitz and Miller, 2008]. If not automated, providing written individual feedback can also be very time consuming.

Right time for individualized feedback

It is not always good to give feedback. In a longitudinal study, various time points are potential candidates for providing feedback: after the first year, after the last year, etc. Feedback might change the awareness of participants with regard to the targeted research question. Then, it would be better to wait and to give feedback at the end of the longitudinal study. The time point in which to provide individualized feedback also lacks research and literature is scarce on this topic. Lefaivre and colleagues [2007] suggest to differentiate between types of studies, especially between cross-sectional and longitudinal. In general, for example in cross-sectional studies, they recommend to wait until the end of the experiment to provide individualized feedback to all participants. In specific settings, as in longitudinal studies with long time intervals between two measurements or in cases where the assessment concerns particularly critical issues, the researcher would need to provide intermediate personalized feedback of some sort. The delimiting line between the two approaches is still unclear and no specific guidelines exist on this topic. It lays in the ethical responsibility of the researcher to design the feedback process including its timeline explicitly.

Appropriateness of feedback

Beside being delivered in the right way at the right time, feedback must also be delivered in the appropriate manner. In the context of individualized feedback, appropriateness combines sensitivity and consideration to show respect for the participants, their dignity and their welfare. This starts by using a vocabulary that laypeople can understand and including both specific and individualized as well as general information. Thus, feedback should clearly state the title of the study, its purpose and its conclusion. Further, appropriateness also requires the supervision of a qualified professional to ensure that the complexity is managed in a way that guarantees the right interpretation. Participants should always be entitled to a written feedback in an easy and understandable language and to ask direct questions about the feedback [Lefaivre et al., 2007].

Framing (undesirable) feedback

Interpreting study results, especially in the context of test or assessments, especially in Psychology and especially if the results are in-

dividualized, is a delicate task that can have deep, unwanted consequences if done improperly [Lefaivre et al., 2007]. For this reason, framing feedback is not easy and requires a lot a thoughtful attention. Pope [1992] recommends to stress the fallibility of the feedback, especially for cross-sectional studies. In cross-sectional studies, the study and thus the entire feedback is based on one day and the contextual factors might be multiple. Even though this variability is not a problem on an aggregate level, it might be very sensitive on an individual level. The sources of bias must be systematically included and discussed in every feedback [Pope, 1992, Lefaivre et al., 2007] and even more if the feedback is in written format since the expert cannot adapt the feedback from the recipient's reaction. However, it was found the participants' negative response to feedback is often to a lesser extent than anticipated by the feedback providers. Pariticipants' ability to cope with undesirable feedback is often underestimated [Shalowitz and Miller, 2008].

Concrete example of undesirable feedback

A concrete example for ethical difficulty tied to individualized feedback is in the case of of children studies where the feedback is to be provided to the parents of the participants rather than the participants themselves. Similar situations could arise in research with older adults (e.g., Alzheimer patients), disabled people, etc. The dilemma extends between the participant's right to privacy and the respect of its right to autonomy and the parents' (or relatives') right to information and moral obligation to advocate for the participant's welfare [Lefaivre et al., 2007]. Finding this balance might not be a problem in the case of desirable feedback but makes delivering undesirable feedback even more challenging.

Normative and comparative data

The question of desirability includes a form a comparison - depending on what is a particular result desirable or not? Researchers and feedback providers should provide feedback on items that are validated and where a high enough degree of consensus exists in the field. This guarantees that individual results can be classified between desirable and undesirable. This requirement for either normative data or empirically-derived thresholds should retain from providing feedback on invalidated items. This ensures that feedback can be described relative to established norms or to percentile ranks that are meaningful. When this holds, the feedback can describe a general explanation on the categorization and what each category means [Lefaivre et al., 2007].

Feedback as compensation

Beside the need for normative and comparative data in individualized

feedback, researchers raised ethical concerns the fact that individualized feedback might be assimilated to research compensation and lead to coercion. Thus, Lefaivre and colleagues [2007] recommend to take this concern seriously and to evaluate whether the individualized feedback could restrain the freedom to consent. To support this caution, they recommend not to advertise about the individualized feedback in the recruitment procedure of participants and to explain in detail the content and the time point of the feedback at the same time as receiving the informed consent. In any case, the feedback must clearly be explained by the researcher to the participant [Lefaivre et al., 2007].

Despite this collection of recommendations and best practices, there is little literature on the best process for giving feedback and on the time and costs linked to providing individualized feedback. Researchers such as Shalowitz and Miller [2008] or Lefaivre and colleagues [2007] call for more research to better understanding what participants' needs are and how to facilitate the generation of individualized feedback.

2.3 Challenges in visualization

An understandable and appropriate way to communicate individualized feedback is by graphically displaying individual values as well as other participants' values, for example in aggregate form. However, visualizing data is challenging and not every graphical representation reaches the goal of providing simple information in a non-misleading way. In this section,

Need for visualization

With the increasing availability of data and the need to see pattern, the field of data visualization has been booming[Few, 2012, Murray, 2017, Munzner, 2014]. Whether it is delivered verbally or in writing, feedback is commonly based on texts expressed in natural language. Per definition, text is linear and does not allow to present the entire shape or any patterns of data at once. Graphs, on the other hand, enable it at the first sight because they are visual and show relationships among the data by allowing them to take a shape [Few, 2012, Murray, 2017]. Graphs are particularly adapted for providing feedback because they might be simple and fast to understand. For ethical reasons, it is not enough that graphs "might be" simple and fast. Every feedback and thus every graph must be appropriate and understanding graphs is often a challenging task. This is why it is important to look deeper at

the elements that help make graphs simple and understandable. At this point, beside the advantages of graphs over texts, it is also worth noting their complementarity. Graphs alone do not suffice to present the whole, complex picture of a feedback and must be accompanied by texts to explain what is seen is the graph and direct the reader to the relevant parts.

Boundaries of visualization

One of the challenges of the data visualization era is the belief that any visualization is good and that everything can be visualized. In her book, Munzner [2014] gives a thorough introduction on what type of data can be visualized. To visualize data, the data must be understood and information gained about this data. The data types can be a rich source of information. She differentiates between five types of data, whose combination form datasets: items, attributes, links, positions and grids. Since this research focuses on application to the field of Psychology, were data are collected in tables, only attributes and items will be further explained. Items, on the one hand, are individual entities, typically discrete, that can represent people, couples or groups. Attributes, on the other hand, are characteristics that can be measured and observed, such as affect, stress or time. Attributes are typically continuous and can be categorical, ordinally ordered or quantitatively ordered [Munzner, 2014]. These classifications helps finding out what can be visualized. With this distinction, it is possible to know what information can be inferred from the data itself and what must be provided as information on top of the data itself.

Wrong use of graph

Knowing what can be visualized is not quite enough to guarantee the ethical imperatives of appropriateness. Data presentation is a topic that has been quite well-studied in theory [Few, 2012, Murray, 2017, Munzner, 2014] but remains underused in practice, also in Psychology. Weissgerber and colleagues [2015] highlight the weaknesses of data presentation in scientific publications and call for a new data presentation paradigm. Their warning concerns primarily the types of graphs and figures that are used to visualize data. With the example of the bar graph, which has been designed for categorical variables and not for continuous data, the raise the concern that poorly chosen graphs can be deeply misleading. When continuous data is represented as a bar chart, they show that many different datasets, with different characteristics and entailing different information, would lead to the exact same bar chart. This is a problem because when information is hidden or lost, it discourages the readers to carefully think about the information displayed. They found out that bar charts suppress distributional information and suggests to use scatter plots or

box plots instead. They also remind that the main goal of visualizing data is to rapidly see patterns and shapes that must be somehow aggregated when the datasets are large. However, for small datasets, they recommend to show every data point without aggregation [Weissgerber et al., 2015]. In the field of Psychology, within-person results are often hidden or lost in the aggregated form of between-person graphs. This is why it is important to differentiate between within-person and between-person graphs.

Comprehension of graph

For guaranteeing appropriateness, beside ensuring that graphs are not misleading by hiding relevant data, the understandability of graphs must be given. Shah and Freedman [2011] investigated the effect of graphical literacy skills and prior knowledge of the reader its ability to correctly interpret the graph. They presented their study participants with graphs of different types and collected data on the participants' general graphical skills and their specific domain knowledge about the data shown in the graph. The output was the interpretation of the main effect in binary form, either well-interpreted of not. They found out that none of the three independent variables alone (graph type, graphical skills and domain knowledge) had a significant effect on the interpretation. Combining the three factors, however, leads to the highest rates of correct interpretation. They explain the effect of the domain knowledge as an incentive to look for trends and ask questions. However, domain knowledge tends to limit this curiosity to expected outcomes. The graphical literacy, on the other side, provide the right skills to search for unexpected results and interpret them correctly. They underline that this would not be possible if the type of the graph would hide or somehow not allow for the correct inferences to be made [Shah and Freedman, 2011].

Data-ink ratio

To objectivize and measure the simplicity and effectiveness of graphs, Few [2012] elaborates on the concept of data-ink ratio. The goal is to highlight the important information and the idea is that in a visualization, not all (potential) ink needed to produce a certain graph directly relates to the data it is representing. Some is used in an indirect manner to support the data by display grids, legends, etc. The data-ink ratio is computed as the percentage of ink used to represent data compared to all ink (direct and indirect) needed. With this concept, Few [2012] emphasizes the need to reduce the indirect use of ink while maximizing its direct use. For the former, he suggests to differentiate between supporting indirect ink and wasted ink. Wasted ink represent elements that do not make the graph worse if they were suppressed. In general, the idea is even that the graph becomes bet-

ter without it and thus must be removed. The supporting indirect ink should be kept but made less prominent, for example by displaying it in the background. For the ink used to directly display data , the differentiation between useful and useless ink should also be made. Not all data responds equally well to the question at hand and some is (for a specific question) just not informative. This data should also be removed because it tends to distract the reader from the important data. The important data should be emphasized so that the reader knows what to focus on (e.g., by making points or lines or bars bigger, brighter or clearer) [Few, 2012].

2.4 Research questions

The topic of giving feedback, in particular individualized and visualized feedback, is an important task for all fields working with human-subjects, including in Psychology. However, in none of these fields has the process of providing feedback become a main research area. On the contrary, it has remained a side topic to which little attention has been paid. The scarce literature and the few studies conducted on this subject all call for more research and experiments in the area of giving feedback.

The present research consolidates the specific need of a group of researchers in Psychology of the University of Zurich to have a tool for providing automated and individualized feedback with the research gaps highlighted in the previous sections. At crossroads between these two aspects, three research questions have been developed and constitute the backbone of this research:

RESEARCH QUESTIONS :

- Research question 1: What are design principles for a feedback creation tool to follow the current practice of giving feedback?
- Research question 2: What are design principles for a feedback creation tool to output individualized feedback?
- Research question 3: How can the visualization for a specific feedback be inferred from the type of data provided?

Chapter 3

Need-finding

In this chapter, the process of need-finding is presented. It explains how the research was conducted and links the research questions with the answers provided by researchers in Psychology. The results from the need-finding phase offer a common mental model for understanding the design decisions that were taken in this research. The indications of this chapter should also enable future developments to pick-up on the work done so far and to continue the path that led to the *Indivi* tool. First, this chapter details the set-up and the realization of the contextual inquiry. Second, it summarizes the results from the user research for each of the three research questions.

Context for the contextual inquiry

To understand where the contextual inquiry started, some context information is helpful. To improve the researcher-participant relationship in time-intensive psychological studies, researchers in Psychology aim at providing additional incentives for participation in form of individualized feedback. As the object of the present user research, these researchers in Psychology will be referred to as "the users". This research started with the users' realization of the amount of work related to providing individualized feedback in large studies involving hundreds of participants. A research project was launched as a collaboration between the University Research Priority Program (URPP) "Dynamics of Healthy Aging" of the University of Zurich and the Zurich People and Computing Lab (ZPAC) of the University of Zurich. The research project was formalized in a brief text reflecting the informal exchanges that had taken place during the upstream preparation. Starting from the high-level expectations provided in the project description, a plan was developed in form of a contextual in-

quiry for finding the needs and eliciting the requirements associated with providing individualized feedback to large number of participants.

3.1 Contextual inquiry

The contextual inquiry consisted in interviewing researchers in Psychology about their view and habits on giving feedback to participants of their studies. One researcher from the URPP served as primary point of contact in the choice and the recruitment of interview partners. As the instigator of the project, she also gave advice and directions whenever multiple alternative existed. She will be referred to as "the principal".

Interviewees profiles

Interviews were conducted with 7 researchers in Psychology. Four were senior researchers (holding a PhD and several years of research practice) and three were junior researchers (PhD candidates). Six were from the University of Zurich, including three from the URPP Dynamics of Health Aging, and one was from the University of Arizona. Three conduct research in Gerontopsychology, two in Clinical Psychology and two in Health Psychology. The types of studies these researchers were currently or had been involved with included cross-sectional studies, longitudinal studies and micro-longitudinal studies. Two researchers were met in a joined interview. The principal was one of the interviewee. The interviews were semi-structured based on an interview grid. The interviews started with a question about their research and whether feedback (individualized or generalized) had been provided to their participants. Participants were asked about their views on feedback (and specifically on individualized feedback) as well as the risks and benefits they see in providing feedback, including the potential influence that feedback might have on participants. Diving more specifically towards the requirements for the feedback tool, participants were asked about the type of data and the kind of variables they need to give feedback about and whether they have preferred types of visualizations. The question of the data preparation and other potential pre-steps was discussed. Finally, the interviewees were asked whether they knew about other or similar tools for providing feedback. Apart from the interview questions, the interviews included showing and commenting feedback artifacts that the interviewees knew about or had produced themselves. During the interview, the researchers were also asked to draw a picture of

their mental representation of a supporting tool for feedback generation. The drawings were then discussed during the interview. The interviewees had no further knowledge about the project than a short introductory e-mail explaining that the interview was part of a user research aiming at providing a tool for individualized feedback generation. The interviews lasted for approximately 45 to 75 minutes. All interviews were audio-recorded and picture of the artifacts and drawings were taken. After each interview, the most prominent learnings were written down in a ”brain-dump” session.

Qualitative analysis of the interviews

All the interviews were qualitatively analyzed by listening to the audio recording and completed by the data from the brain-dump session. Each piece of information was written down using yellow post-its. Questions and unclear statements were written on orange post-its. Pieces of information or questions that appeared multiple times (either in the same interview or across interviews) were written down once. From this process, more than 200 pieces of information were collected on yellow post-its and around 15 questions and unclear statements were collected on orange post-its. In reviewing the data, similar pieces of information were clustered together. After all post-its had been placed together, the resulting clusters were named. This exercise resulted in an affinity diagram (Figure 3.1) classifying the data in 14 topics:

- Types of studies
- Types of data
- Types of results
- Types of feedback
- Best practices in providing feedback
- Scope of the feedback (width and depth)
- Time for feedback
- Feedback output
- Types of visualizations
- Layout wishes
- Generation process / configuration of feedback
- Identification to receive feedback
- Benchmarks
- Gamification

Figure 3.1: Affinity diagram

Derived requirements

From the affinity diagram, requirements on the final software were formulated and classified into three categories: ”the software has to...”, ”the software should...”, ”the software could...”. The translation from the collected data in concrete requirements resulted in 80 requirements, 49 in the category ”has to”, 13 in the category ”should” and 18 in the category ”could”.

Follow-up interviews

From the data gathered during the interviews and evaluated with the affinity diagram, the exact process for generating and providing feedback according to current practices could not be precisely drawn. Follow-up interviews were conducted to better understand the concrete process used by researchers to provide manual or semi-manual feedback. Three researchers participated in the follow-up interviews, all from the University of Zurich. Two were senior researchers and one was a junior researcher. Two were from the URPP Dynamics of Healthy Aging (researchers in Gerontopsychology) and one was Health Psychologist. All had at least some experience with providing feedback, two had experience in providing personalized feedback. The principal was one of the interviewee. During these interviews, the interviewer and the interviewee worked together to conceptualized and visualize the personal process of the interviewee (i.e. the steps that they had or would actually take). While the interviewee was thinking at loud, the interviewer was writing down the steps on post-its (one step per post-it). Once all steps had been named and placed in their chronological order following the current practice, the interviewer and the interviewee discussed whether the steps could be rearranged to be more fluid and avoid repetitions. The interviewee

was then asked which part had to be made by the researcher himself and which parts could be automated. The resulting process was photographed for future analysis and the post-its were collected. The processes from the three interviews were compared and aggregated. Data collected during the first interview round was also taken into account to produce an "ideal process". The "ideal process" was visualized to facilitate the communication with the principal and other interviewees in further steps of the process.

Communicating ideas with prototypes

To materialize ideas, communicate mental models and provide support for discussion and testing feedback, prototypes were developed. They took the form of paper prototypes, story boards and finally high-resolution, coded prototypes. Prototypes were developed throughout the need-finding process. They have been primarily used in the communication with the principal to validate or discuss possible options. The first prototypes were generated by the users themselves during the first round of interviews. The explained their mental model of the future tool with means of a paper prototype that was then used as a basis for the discussion. From all the paper prototypes, a general idea emerged with respect to the format and the flow of the tool as well as its main components. From this prototype, concerns emerged regarding the exact process of the feedback generation and led to the follow-up interviews.

Iterations of prototypes

After having refined the understanding of the current process, a more detailed prototype show-casing various options was created. It served as basis for discussions with the principal and made the problem of the data classification and organization be clearer and more tangible. To compare ideas on possible structures and classification of data to automate the generation of visualizations, each possible constellation of variable types and data format was developed on its own and organized as a visual tree, which prototype served as basis for discussions and agreements. Finally, prototypes were also developed to share the mental representation of the final output with users. This allowed to conduct discussions on the structure, format and style of the output. After having reached a high enough level of understanding about the requirements on the final tool, paper prototypes were replaced with software pieces that featured the prototype much more closer to what it should look like in the end. The high-level prototypes switched the focus of the discussions less on the logic and the structure and more the user experience. Using prototypes along the way throughout the entire need-finding phase helped make ideas tangible, supported the communication with the uses and enabled the crystallization of prob-

lems that then led to improvements.

3.2 Results from need-finding

3.2.1 Feedback process

Results regarding the first research question

The first research question aimed at finding out how a feedback creation tool should be built to follow the way researchers give feedback today. The goal was to create a tool that is close enough to current practices so that it is familiar to use for experienced researchers and intuitive for researchers having no experience in providing feedback. The results from the interviews showed quite some variance in the way researchers provide or expect to provide feedback to their participants. Although researchers who previously had been involved in providing feedback to participants all mentioned some steps during the interview, none had a very clear idea of how it should be done and what steps were involved. From the interviews, it became clear that

- no common denominator / best practice existed on how feedback should be given to participants,
- junior researchers were not taught how to provide feedback to participants,
- there was some confusion about providing feedback to the participants in a manner that was understandable by them and simply send the participants the results of the research as it had been prepared for peer-reviewed publications,
- the process of giving feedback was only rarely orientated towards giving participants an individualized feedback.

Process representation

The lack of clarity on the common denominator imposed to have some follow-up interviews as described in the previous section. Even during these interviews, the steps and their chronological order had to be worked out in a co-creative manner with the interviewer. This lack of clarity among experienced researchers explained why junior researchers could not be clearly taught about the process. The confusion about the type could be worked out with the principal and a clearer focus could be communicated during the following interviews. The

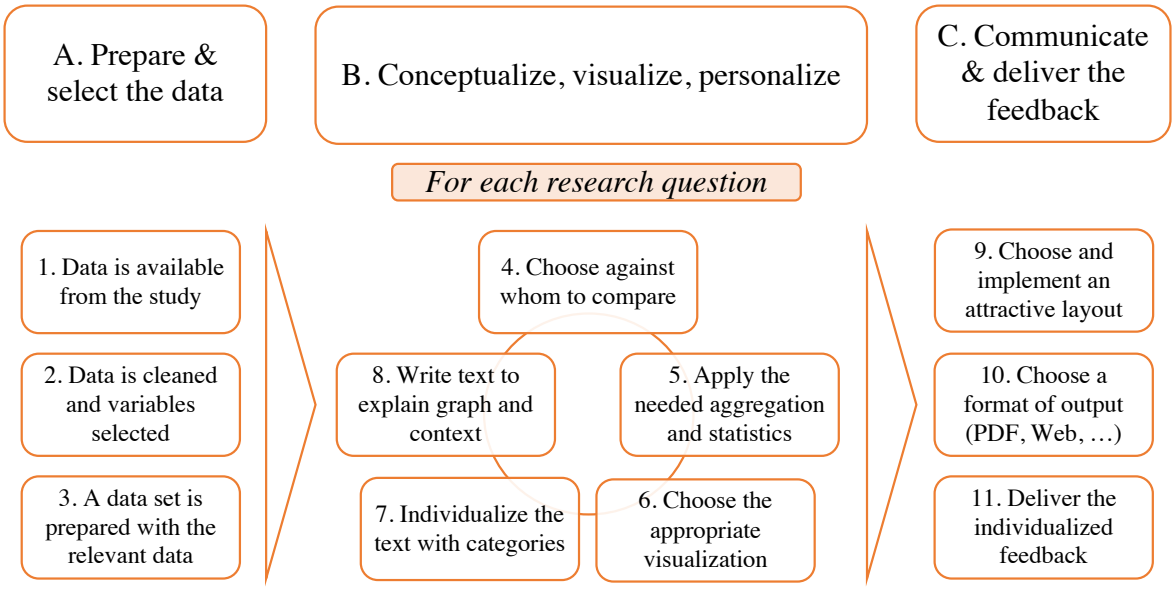


Figure 3.2: Common denominator of current processes for providing feedback

orientation of the process towards the individual participants (rather than the research focus) helped to lay the ground for a first common layer: the feedback had to be constructed in a way that was logical for the participant independently of the type of research conducted by the researcher providing the feedback. Despite the variation, after the second round of interviews, a tentative process could be developed and was validated with the principal (Figure 3.2)

Feedback preparation

Figure 3.2 shows a process in three blocks (A, B, C) represented by the pink post-its and detailed in 11 process steps represented by the yellow post-its above them. A) The first block is about preparing and selecting the data. The process starts with step one: the researchers gathers data within their research and prepares the data for feedback. Second, the data is then cleaned and the relevant variables for the feedback are selected. Third, the data relevant for the feedback is summarized into a table or a data frame. This ends the block of preparing and selecting the data.

Feedback generation

The second block in the feedback process is to conceptualize, visualize and personalize each research question on which feedback should be provided. Thus, this block is iterative and can be undergone as many times as they are research questions for which feedback is wanted. The five steps included in this block can be placed in any order as they all

depend on each other. The order in which they are presented here is a possible logical combination of them. For each research question, the researcher has to define if the feedback should be given as a between-person or a within-person comparison. From a participant's perspective, this means either seeing his values compared to the values of other participants or seeing his own value over time compared to his personal mean. The researcher also has to decide on the aggregate statistics to use for the comparison, being the mean, the standard deviation, the sum, the median or any other percentile measure or even the of distribution. Each research question used for feedback should be visually represented in form of a plot reflecting the choices above. In order to individualize the feedback while avoiding writing a text for each single participant, the researcher can define categories of participants based on the participant's values. For each research question, for each visualization and for each category defined, the researcher has to write an explanatory text in a language understandable by the participant. This ends the block of conceptualizing, visualizing and personalizing each of the research questions in the feedback.

Feedback delivery

The third block is about communicating and delivering the feedback and has to be done once after all research questions have been defined. The first step is to choose and implement an attractive layout. The second step is to choose an output format, for example as a individualized website, a printed booklet or a PDF document. The third and final step of the feedback process is to deliver the feedback to the participants, for example via e-mail, via postal services or during social events and presentations.

On the general activity of giving feedback, the need-finding from interviews resulted after an iterative process in a consensual description of the way researchers in Psychology provide feedback today. The three main phases are the preparation of the data, the generation of the feedback and the delivery of the feedback.

3.2.2 Individualization

Results regarding the second research question

The second research question was about individualizing the feedback and outputting an individualized feedback. The goal was to go away from a practice where all participants received the exact same document as a feedback independently of their own values and these of the group. The general idea underlying this research was that participants

are interested in their own values and that receiving their own values represents an incentive for participating and remaining in a study. The results from the interviews gave insights on three approaches on how to individualized feedback:

- Individualize graphs
- Individualize texts with placeholders
- Individualizing texts by categorizing participants

Individualized graphs

The most straight forward idea that came to mind during the interviews, supported by the possibility of technical support, was the generation of graphs and plots displaying at least the values of each participant individually and potentially more information. However, it was stressed that visualizations should never be delivered alone but always accompanied by texts. From this insight, interviewees provided ideas on how to individualize texts.

Individualized texts with placeholders

In practice, researchers reported to write multiple texts on each research question. Often, feedback texts are written in three parts: a general explanation and motivation of the research question, an explanation about the graph and an interpretation of the graph highlighting the results displayed. The need for individualization for the first two categories is weaker as they are more general. However, researchers expressed the need to insert placeholders when writing these texts. These placeholders should refer to either a summary statistic of all participants (e.g., the group mean, the individual value or the individual average) for a participant.

Categories of participants

To individualize the third text block, the one interpreting the graph, researchers proposed in the interviews to classify the participants in categories and to write a text for each category. The participant would see the text corresponding to his own category. On the possible categories, researchers most mentioned a classification of type ”high”, ”medium”, ”low” depending on the research question. A possible way of computing these three categories emerged to be the percentile rank of a participant compared to the other participants, where participants above the 75th percentile would be considered ”high”, the participants below the 75th percentile would be considered ”low” and the ones in between the 25th and the 75th percentile would be considered ”medium”. In the case of time varying values, an experienced senior researcher proposed to use the metric of mean square successive differences (MSSD).

Output format of individualized feedback

On the output format, the result from the interviews were very diverse with regard to potential and new ideas that emerged from the discussion but the formats actually used in practice were more restrained. The latter category included regular PDF documents, booklets that were printed either for all participant or with some individualization and a personal static website in the case of one study. The ideas on how feedback could be delivered ranged from creating (even individualized) videos, chatting with a bot for the participants to be able to express their wishes, to ask questions and to give feedback on the feedback or to invite their own health professionals to see or exchange on the feedback. In the current practice, the feedback was distributed via postal services, via e-mail or during presentations and information events. For clinical studies, the feedback was sometimes delivered through the primary care physician. Further, interviewees mentioned the importance to not only give individualized feedback but also to give this feedback in person, especially if the content of the feedback is critical. In general, an important feature of feedback is to be written in "participant-language", meaning using words and concepts that are familiar to participants and refraining from using research-related jargon.

On the individualization of feedback, the need-finding from interviews resulted in three leads: individualizing graphs, individualizing texts with the use of place holders and writing different texts for different participant categories. Insights were gained also with regard to the output format of individualized feedback either as PDF document for electronic delivery, as booklet for print, as presentation for participants' groups or as websites.

3.2.3 Data visualization

Results regarding the third research question

The third research question concerned the link between the data needed for the feedback and the visualizations. The interviews showed that data visualization was a central part in providing individualized feedback. Interviewees reported various types of data including questionnaires, results from cognition tests, epigenetic data, biological measurements (body weight, body height, hormones, blood pressure), sensor-data such as GPS or accelerometer data as well as socio-demographic aspects but also non-numeric data such as audio files or brain images. Interviewees also mentioned ways classifying data, for example between-person and within person values, dyadic

data, fluctuations and joint fluctuations, real-life and lab-data. Researchers reported using this data sometimes separately, sometimes jointly. However, it came out from the interviews that there were very little or no guidelines existed on what makes a good visualization and in what way data should be visualized. The materials gained during the interviews still provided valuable insights that could be exploited to classify the data types.

Data type classification

As for the process of giving feedback, where there was also little clarity among the researchers, structuring the problem from the participants' perspective helped. In the present case of classifying the data, the question could be reformulated as: "What are typical research questions for which feedback is given and what data is needed to plot a graph?"

Intensive work with the principal and another researcher lead to the hypothesis that a tool could provide a visualization for any research question given the following data elements:

- A unique identification for each participant
- A main variable of interest
- Possibly a second and/or a third variable to compare with
- Possibly a time indication (such as days, hours, time points)
- Possibly a group indication (e.g to compare young and old) or a dyad indication (joint response of two participants)

Combination of data types in a tree

These elements could then be combined in a tree displaying all the possible combinations of elements needed to provide a visualization (Figure 3.3). These 8 types of data became the basis for creating prototypes of data visualization depending on the data provided in each type. Figure 3.4 shows a prototype displaying the visualization for all possible combinations of elements needed to provide a visualization. This prototype was created to validate the hypothesis of the data elements needed to give feedback on possibly every research question involving numeric data. The prototype consists of nine parts, displayed in three columns and three rows. The first part shows the general logical while the other 8 show each combination that emerged from the data tree. For each combination, beside the visualization itself, the prototype shows different options, refining questions to the user and a space to enter text according to categories. On the top right corner

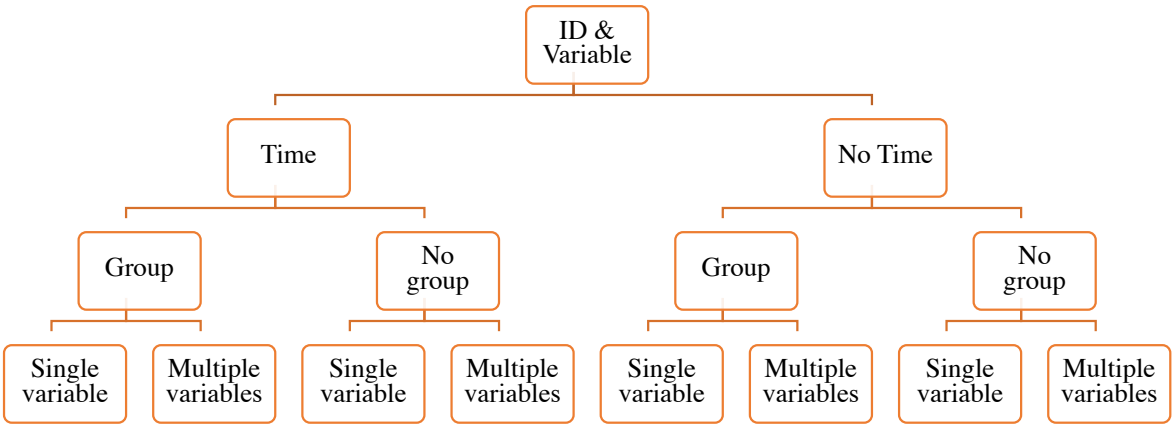


Figure 3.3: Tree with all possible combinations of elements needed to provide a visualization

of each of the 8 screens, a table shows how the data set should be organize to provide the right data for this combination. This was added for the sake of facilitating the communication with the users but was not intended to be part of the final layout.

Types of plots

After the interviews, because of the lack of previous knowledge of the good practices in data visualization, it has been decided to let the tool impose the graph type to the researchers and a prototype was developed to show what combination of data would result in what type of graph. It was found that the types of visualization could be split between the research questions containing a time component and the ones who do not. For the research questions with a time component, the best graph type turned out to be the line plot showing the time on the x-axis and the scale of the variable of the y-axis. This kind of line plots allows showing the development throughout time for a single participant, which is the goal of showing within-person differences. For the research questions with no time component, the focus lays on the comparison with others in the group. To show the distribution of the values throughout all participants, box plots turned out to be the best suited type of graph. Within the box plot, the data point of a particular individual can be highlighted, allowing for a seamless comparison with the group (between-persons). Having the individual and the distributional information within one plot allows to display multiple box plots next to each other, for example in the case where multiple variables need to be compared with each other.

Further validation for future development: conduct a workshop

In order to further validate the finding from the interviews and to gather feedback about the prototypes, the idea emerged of organiz-

Figure 3.4: Prototype of the visualization for all possible combinations of elements needed to provide a visualization

ing a workshop with the researchers that had been involved in the process. The goal of the workshop would have been to verify the common denominator developed, discuss the form and layout of the output and validate the types and visualizations of data taken to be processed. The result of the workshop would have been a clear list of requirements in three categories: "the software has to...", "the software should...", "the software could..." so that all would know what to expect. This would have enabled to focus the testing feedback to matters that were included in the project. Because of the lack of users' availability, such a workshop could not take place.

On the visualization of data, the need-finding resulted in the classification of data types in a tree according to whether the research question on which the feedback was given included a time component, a group component and multiple components. From the resulting eight categories, different types of plots could be inferred. Two main groups of plots were developed: box plots for research questions without a time

component and line plots for research questions with a time component.

Chapter 4

Design decisions for the implementation

One of the main goal of the present research was to design a tool to support researchers in Psychology for them to create individualized feedback for the participants of their study. After an extensive phase of need-finding based on context inquiry, a general design for the tool emerged in a common mental model of the developer and the users as well as in form of diverse artifacts such as an affinity diagram, a visualized process, a data tree and several prototypes. The tool developed was called *Indivi*. While the technical implementation of *Indivi*is discussed in the next chapter, this chapter focuses on the design decisions that were taken for the implementation and links implemented elements with the results of the need-finding. The presentation of the design decisions follow the three research question: how the process of giving feedback was implemented in the flow of the *Indivi* Web-App, what elements were implemented to facilitate the individualization and how the visualization of the data was inferred from user-input.

4.1 *Indivi* workflow

The first research question concerned the design principles needed for a feedback creation tool that follows the current practice of giving feedback. The results from the need finding phase showed that the current practice of giving feedback was not very clear and that no

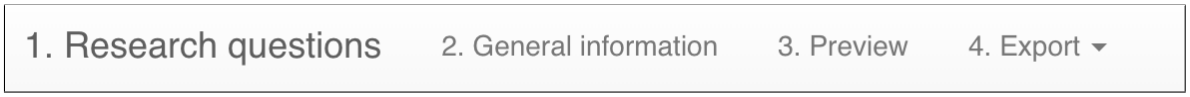


Figure 4.1: The flow of the *Indivi* visualized in the navigation bar

common denominator existed at first sight between the practices of the interviewed researchers. After a second round of interview with a co-creation approach, a process could be designed that reflected the common denominator of all practices. This process has been described in the previous chapter.

Indivi follows the flow of the current feedback practice

The first design decision for the implementation was to use this process to model the flow of *Indivi*. To make the flow of *Indivi* clear to users and available at any point in time, the process steps were visualized and implemented as routing links in the navigation bar (Figure 4.1).

Preparation of the data: upload a file in *Indivi*

The first block of the process, preparing and selecting the data, was decided to lay within the responsibility of the researcher. The tool kicks in after a table has been prepared following to the instructions provided on the website. Since this step is very simple and requires only two clicks, it was decided that it should not be a page on its own but be included on the first page. To increase the readability of the users (especially first time users), collapsible panels were used to display instructions and information on the data format if the user has not yet uploaded data. After the upload, these panels close and information on the feedback generating process is displayed.

Generating feedback in *Indivi*

The second block of the process developed in the need-finding phase is the conceptualization, visualization and individualization of the feedback. This block is the most central in the process, the one that can benefit most of technical support and also the most challenging one. For these reasons, this block has been separated in two pages, one on the conceptualization and one on the visualization and the individualization.

Definition of research questions in *Indivi*

The conceptualization of the feedback corresponds to the definition of research questions about which feedback should be given. The first page of the tool, beside the data upload functionality described above, enables the definition of research questions. Defining a research question for the feedback means for the researcher to provide a name for the research question and to select the data that is needed to create a

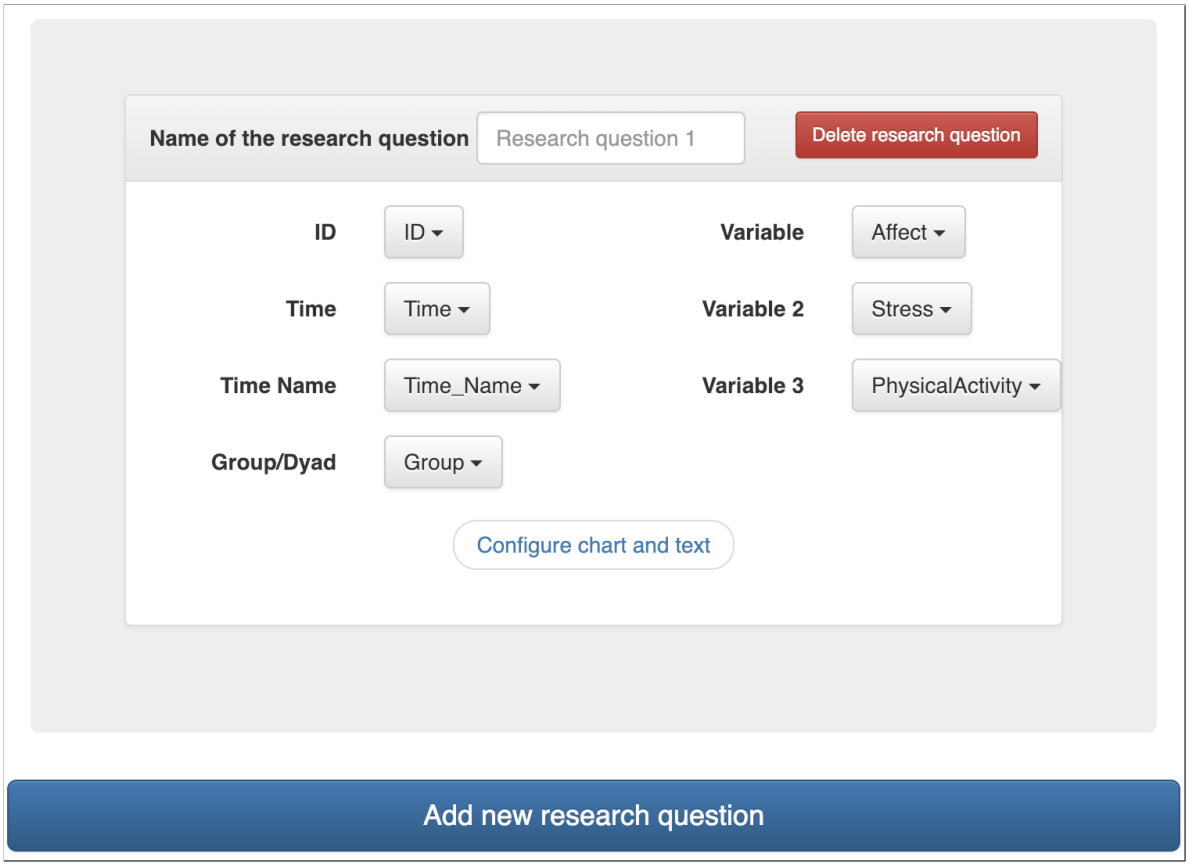


Figure 4.2: The interface for defining research questions

visual representation of the research question. Figure 4.2 shows the interface implemented for defining the research questions. This interface allows the user to define both the breadth and the depth of the feedback. The breadth is defined as the number of research questions to be included in the feedback. The user can use the buttons "add new research questions" and "delete research question" to increase or restrain the scope of the feedback. The depth of feedback is defined as the quantity and the type of data to include in each research question. Following the data tree that resulted from the need-finding phase, the user can select the following data:

- ID: a unique value to identify each single person or dyad (mandatory, e.g., *PartId*)
- Variable: the main value measured for each participant (dependent or independent variable) (mandatory, e.g. *Affect*)

- Time: a numerical value for which the variable varies within each ID (e.g., *Days*, such as *1, 2, 3*)
- TimeName: a textual value for the name of the time points (e.g., *DayName* such as *Monday, Tuesday, Wednesday* or *Jan 1st, Jan 2nd, Jan 3rd*
- Group or dyad : an information relative to each participant ID (e.g., *male/female* or *younger/older*). If only comparison to the entire sample are expected, leave blank. For dyads, "ID" must be the ID of the dyad (*dyadID*) and "group" must be the distinguishable feature (e.g., *husband/wife*)
- Variable 2: a second variable to compare to the main variable (e.g., *Stress*)
- Variable 3: a third variable to compare to the main variable (e.g., *Physical Activity*)

Configure the research questions in *Indivi*

For each data type, the dropdown button displays all column names from the uploaded data. With this interface, the user completed the conceptualization phase of the feedback creation. The visualization and the individualization task happen in an iterative fashion for each defined research question when the user clicks on the button "Configure chart and text", which leads the user to a new page where the research question can be configured. As the visualization and the individualization are central to this research, their concrete implementation will be described in the following sections while this section focuses on the general flow of *Indivi*.

Navigate between the research questions

At the bottom of each research question's configuration page, buttons as shown in Figure 4.3 allow the user to navigate both between the research questions and the feedback process steps. If the user is currently configuring the first research question, the top-left button "Configure previous research question" is not displayed. If the users is configuring the last research question, the top-right button "Configure next research question" is replaced by the button "Add new research question". The bottom-left button "Overview of the research question" leads the user back to the first page where the conceptualization of the research questions took place. The bottom-right button "Go to study information" leads to a new page

Wrapping up the feedback in *Indivi*

As final touch to the feedback, the user has the opportunity to wrap up all the research questions in a general introductory text appearing before the first research question and a concluding text displayed after the last research question. Further, the user can provide a logo of his

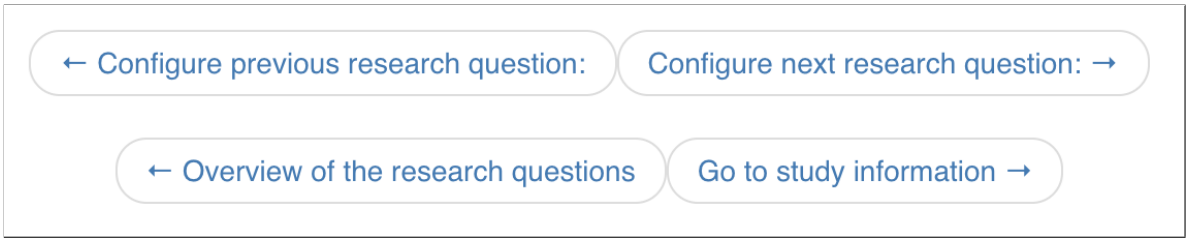


Figure 4.3: Navigation both between the research questions and the feedback process steps

organization, a general title for the feedback and a signature, e.g. his name or the name of the research team. This information can be input by the user in a classical form available after clicking the button "Go to study information" in Figure 4.3 or from the "General information" button in the navigation bar (Figure 4.1)

Previewing and exporting the feedback in *Indivi*

The third block of the process is about communicating and delivering the feedback. It has been decided that the delivery step would entirely lay under the responsibility of the researcher. To facilitate the distribution, the tool outputs one document for each participant. The documents can be downloaded separately for each or participants or all at once with a button "Download all participant feedbacks". To allow for a flexible distribution of the feedback by the researchers, three output formats have been implemented according to the results of the need-finding phase:

- a traditional A4 PDF document best suited for distribution by e-mail
- a A5 double-sided PDF that is best suited as printed booklet to be sent by postal services and handed in during an in-person event
- a static HTML document with the feedback, that can be made available online to the participants, e.g. with a personalized (and protected) link

For convenience and because of the computation resources needed to generate all feedbacks, especially when there are many, the tool provides a "Preview" option. The preview display for one random participant the feedback in all three formats. The preview is available after clicking the button "Preview of the feedback" on the "General information" page or from the "Preview" button in the navigation bar (Figure 4.1).

4.2 Individualizing feedback in *Indivi*

The second research question concerned the ways to individualize the feedback provided to the participants, which are assumed to be more interested in their own values than in the general results. From the need-finding, three approaches emerged for individualizing the feedback: individualizing graphs, individualizing texts with placeholders and individualizing texts by categorizing participants. Their implementation in the *Indivi* tool is presented here. As mentioned in the previous section, the steps of visualizing and individualizing the feedback are done iteratively. For each research question, each participant wants to know how he performed during the assessment. This can be achieved through graphs, as mean of visualization, or through texts which display the values of each participant. Beside receiving their own values, participants are interested and ethically entitled to receive an interpretation of the results in a language that is understandable. The highest level of individualization would be reached if each participant would receive an own interpretation but the practicality is limited both by the resources of the researchers and the number of participants in a study. In some cases, participants are interested in the comparison to others in their group or more generally to all others in the study. Further, results from the interviews stressed the link between the individualization of the feedback and the different formats of output.

Individualized graphs in *Indivi*

The first decision to enable the individualization of the feedback was that the feedback on each research question would contain a graph and that each graph would display the own value of the participant in form of well labeled points. The graphs must have the possibility to display distributional information where the participant sees his value in the distribution as a comparison to others. Others must be differentiated between a specific group each participants belongs to (e.g., male versus female, young versus old) or a general comparison to all other participants of the study. For some research questions, more than one value is of interest. It was decided that the graphs need to account for up to three variables that have to be displayed close enough to each other to allow for comparison between the multiple variables. If answers where given overtime, the time aspect should be visible and the focus change from a between-person to a within-person interpretation, meaning that the comparison to others must be less prominent than the variation over time of the particular individual. However, comparison the group can be given by showing the

group mean and/or its standard deviation as a constant on the graph. These requirements on the individualized graphs follow directly from the data tree (Figure 3.3) that emerged from the need-finding and was explained in the previous chapter. In the next session, each possible combination will be presented in detail.

Individualized texts in *Indivi*

The second decision for individualizing the feedback was to implement a way to display individual values in texts. The chosen approach was to create placeholders that were made available to the participants as a list in the instructions. The placeholders were formatted with a hashtag (#) as the first character so the program could search for them using regular expression and replacing them by the value. Once these placeholders were implemented, they could be enhanced by adding other placeholders on general statistics over the sample or a group, such as sample mean or sample variance. Providing a list of which placeholders are available instead of allowing free text allowed to define stable variables in the code that referred to the desired value. However, in order to make the list easily readable and not overwhelming, the variety of offered placeholders was kept small, simultaneously reducing the researcher's freedom in writing his feedback. The constraint was nevertheless considered valuable to limit the variation in the types of feedback and forcing researchers to provide simple feedback. To increase the flexibility, the available placeholders depended on the type of graph displayed, which depended on the types of data chosen according to the data tree (Figure 3.3). Placeholders were provided for both the mean of the individual (person mean) and the mean of the sample (sample mean). Figure 4.4 shows the instructions on and the list of available computed values in the case where three variables were selected (”Affect”, ”Physical Activity”, ”Stress”). Since the program computes the person mean and the sample mean for each variable, this combination results in six computed values. Below the list, the use of these computed values in text are showcased. Where number appear, the corresponding ”tag” was typed.

Text categories in *Indivi*

The third decision for making the feedback more individualized was to enable the researcher to create different texts that would be displaying to the participant according to their values. It was decided to implement three categories of participants following the scheme ”high”, ”medium”, ”low”. Participants with a score above the 75th percentile were considered of the category ”high”, participants with a score below the 25th percentile were considered of the category ”low” and participants with a score between the 25th and the 75th percentile

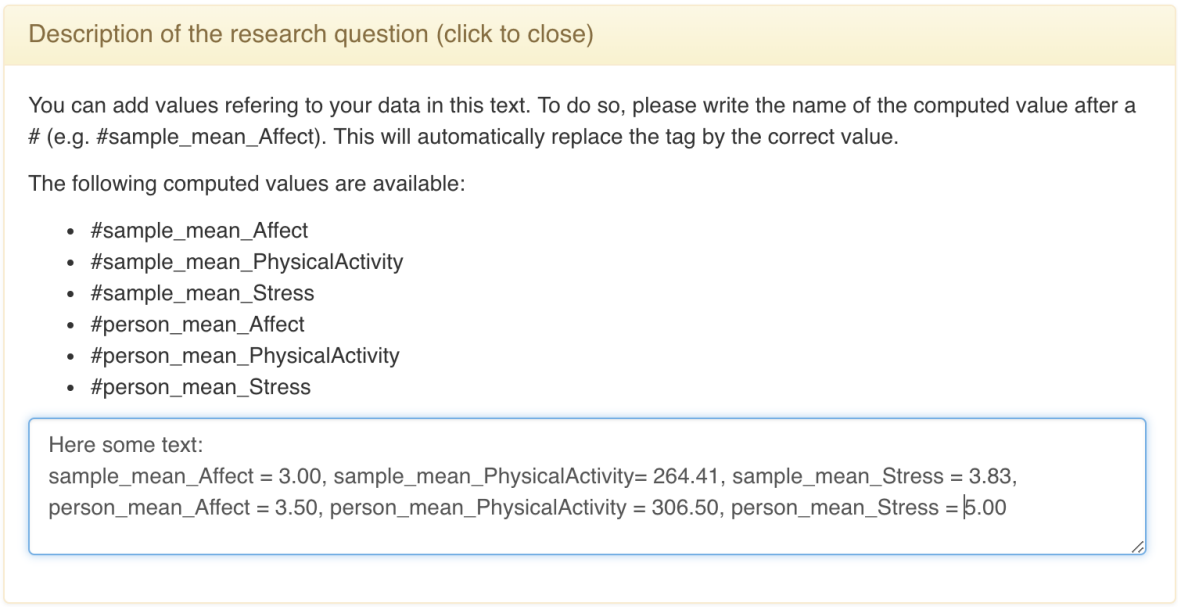


Figure 4.4: Instructions on and list of available computed values and their use in text

were considered of the category ”medium”. If more than one value was available for each participant, the mean of the values was computed, giving a score that allowed for comparison. When each participant had multiple values over time, another score was computed instead of the arithmetic mean: the mean square successive difference (MSSD) which was proposed as a variability measure over time that better captures the ups and downs [Jahng et al., 2008, Houben et al., 2015]. The MSSD also has the advantage of giving one score for each participant, allowing for the computation of percentiles and the classification in categories. In the case of multiple variables, the main variable, for which the ranking was computed, was the variable given under ”Variable” (as opposed to the ones given under ”Variable2” and Variable3” in the conceptualization phase. When generating the individual feedback, for each research question, the program would first compute the category of the participant based on his own data and the data of the group and then display the appropriate text. Figure 4.5 shows the instructions and the text fields that the research has to fill so that each participant will receive a text according to its own results. Again, these texts can display the person and the sample mean using the tags described above.

The individualization of the feedback in the *Indivi* tool takes place by providing a graph to each research question which clearly shows the

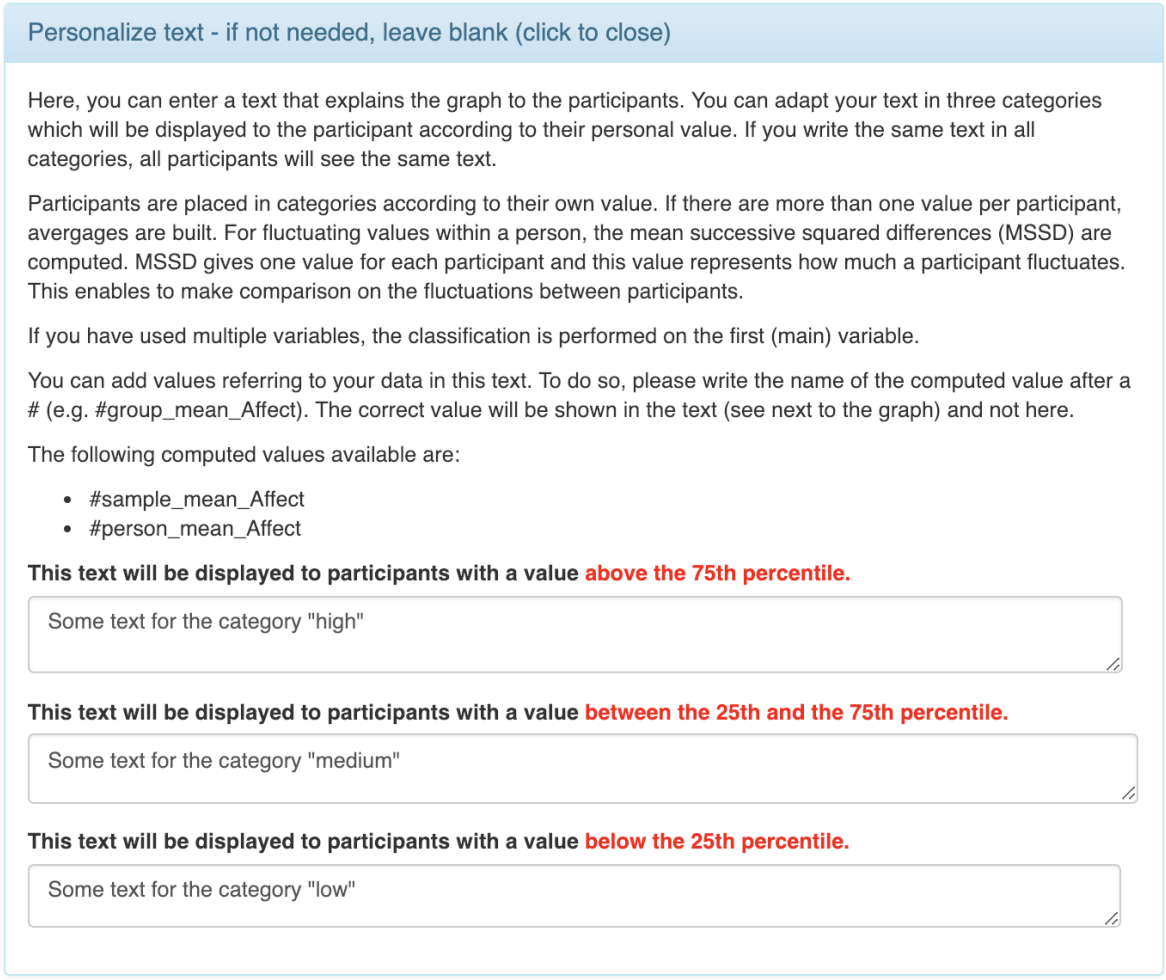


Figure 4.5: Instructions and fields for texts specific to a participant's category

individual value, by allowing the use of the individual value as well as the sample value as placeholders in the texts and in computing three categories of participants based on their individual values for which each category will receive a text adapted to their score. With the individualization came the need for an output format adapted to the delivery and three types of output were chosen: a classical PDF document to be attached in an e-mail, a printed booklet to be sent to the participants via postal services or personally given and a HTML static website that could be regularly updated by the researcher and therefore useful for participants in longitudinal studies with long time periods between two measurements.

4.3 Visualization of data by type

The third research question concerned the visualization of individualized data. Having a clear understanding of the process and after implementing ways to individualize the feedback, the main challenge of this task came from the fact that there is no "right way" of visualizing data and that many possibilities exist. Following the insights gained in the literature (Section 2) and the results of the need-finding phase summarized in the data tree (Figure 3.3), different graphs are generated by *Indivi* depending on the type of variable included in the preparation of the research question.

Types of graphs

Each possible combination results in eight different types with some types being further differentiated in their visualization as in the case of groups, where the user can chose to display a comparison between groups or within groups, or when the group is a dyad and the individual results of both dyad members must by displayed. To ensure continuity and to avoid overwhelming participants with too many forms of graph, two main graph forms were chosen: the box plot and the line plot. The box plot shows the distribution of all answers and offers a good but static understanding of how other participants performed and where the individual is located in this distribution. Box plots are not very illustrative when an item is measured over time. When a time variable is included, *Indivi* displays a line plot where the x-axis is the time line and the y-axis the value of the measured item. In this case, the distributional information is lost but replaced by the dynamics over time. The individual raw values are displayed while the values of other are averaged and the mean at each time point is shown.

Type 1: just one variable

The simplest combination of data consists of one variable with values for each participant, no group and no time, when the user chooses only the *ID* and the *Variable*. The data is displayed within a simple box plot (Figure 4.6) where 50% of the participants have a value within the box, 25% above the box and 25% below the box. The line in the box shows the median value (i.e. 50% of the participants lay above and 50% below). The whiskers show the maximum and the minimum value given by any participant. The black square represents the value of the participant seeing the plot. This way, the participant gains information on two elements: first on his or her own value and second on how this value compares to other participants' value. The value showed in Figure 4.6 can be interpreted as an individual value of 3.5,

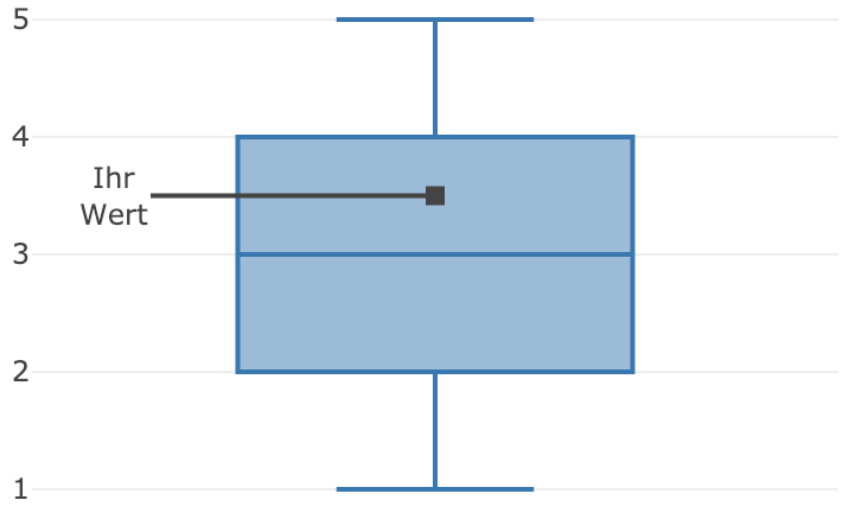


Figure 4.6: Type 1: just one variable

which places the participant in the higher half, the third quartile and within the typical range (the box).

Type 2: adding a group component

The second type differs from the first type because it includes a group component on top of the variable of interest, where the user selected data for *ID*, *Variable* and *Group*. As mentioned above, adding a group can have different purposes which have an impact on how the information needs to be displayed.

Type 2a: comparing between groups

A reason why a group component is added is to compare the values between two distinct groups of participants. In Figure 4.7, the two groups are *male* and *female* and the graph displays the value for a male participant. The participant gains three pieces of information: first about his own value, second about how he does compared to other participants in his group and third how his group's values compare to the other group's values. In this particular case, the participant has an individual value of 3.5, belongs to the typical range of male participants, which is higher than for female participants.

Type 2b: comparing within groups

In some cases, it is not relevant to display the comparison between groups to the participant but it is still relevant to compare each participant to his or her group and not to all participants. In Figure 4.8,

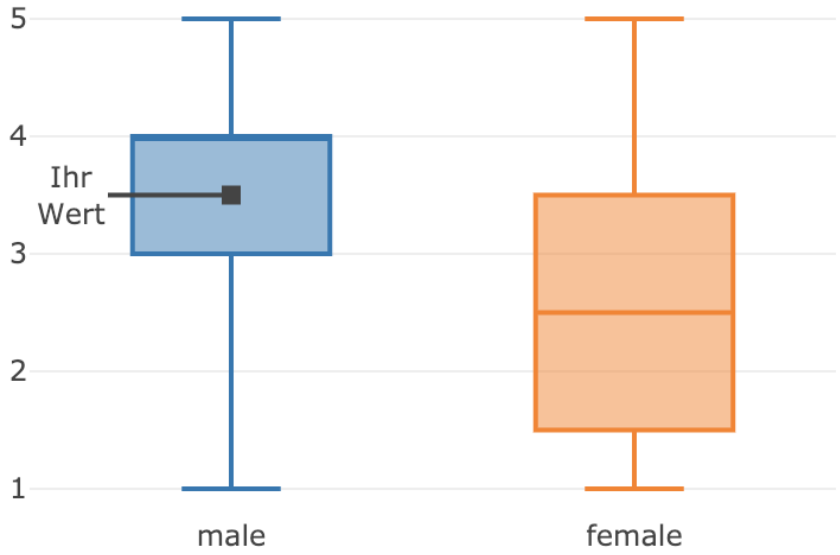


Figure 4.7: Type 2a: Comparing between groups

the participant belongs to the group *male* and gains information about his own value (3.5) as well as about how he compares to other participants in his group, in the typical range of male participants (within the box). In the case of this particular data, the median is equal to the 75th percentile (the upper limit of the box), meaning that the entire third quartile had a value of exactly 4.

Type 2c: Dyad as group

In the dyadic research, the participants belong to small groups of size two (the dyad) where the each member has a distinguishable feature (e.g., a patient and a doctor or a male and a female in an heterosexual couple). For dyads, the user provides the ID of the dyad under *ID* and the distinguishable feature under *Group*. In the visualization, the value of each member must be displayed. In Figure 4.9, the dyad is a heterosexual couple and both the values of the male and the female individuals are displayed within the distribution of their own group. The participant gains four pieces of information: first his own value (3.5), second the value of his dyadic counterpart (2.3), how he does compared to others in his group and how his counterpart compares to others in her group.

Type 3: Multiple variables

The third type is an extension of the first type. Here, multiple variables are compared with each other (without the group component).

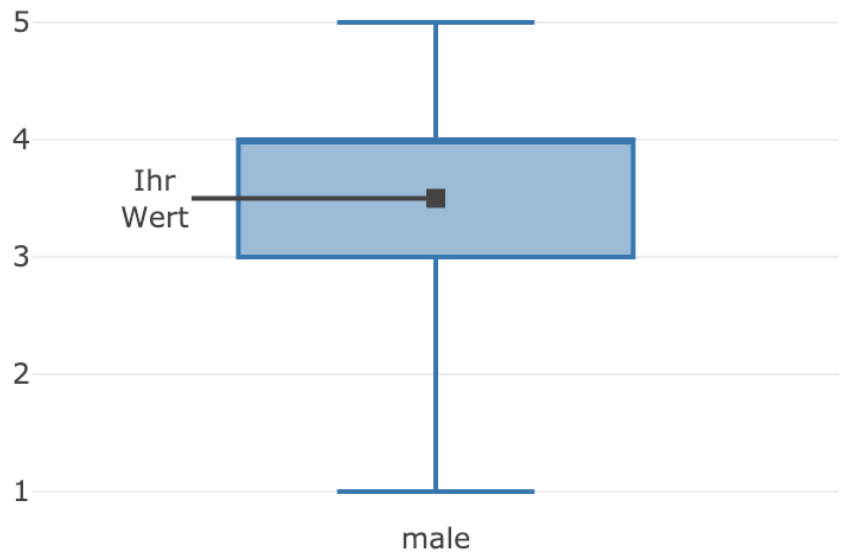


Figure 4.8: Type 2b: Comparing to a subgroup

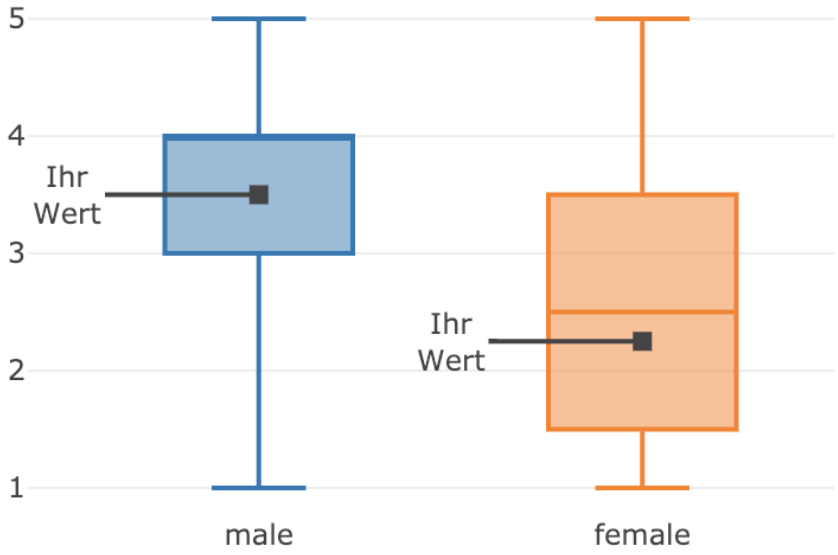


Figure 4.9: Type 2c: Dyad as group

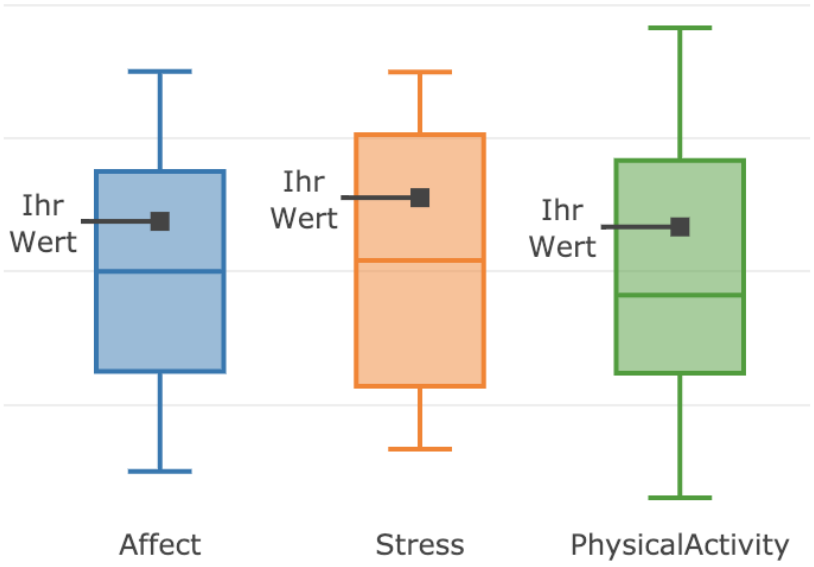


Figure 4.10: Type 3: Multiple variables

Indivi supports comparison of two or three variables but not more as it would make the graph too crowded and too difficult to read. However, it is still possible to add new research questions to compare more variables. Since the variables are mostly not measured on the same scale, the values are centered using the z-transformation and the axis values are removed because the transformed values do not have an intrinsic meaning anymore and to avoid confusion. In Figure 4.10, the participant gains four types of information: first, how he does compared to the other participants regarding his *affect*, second regarding *stress* and third regarding *physical activity*. For all three items, the participant is in the third quartile (i.e. in the higher part of the typical range). Fourth, the participant can see how his relative value compared to others differs for the three items *affect*, *stress* and *physical activity*. In Figure 4.10, his relative score is very similar for all three items.

Type 4: Multiple variables and group component

The fourth type differs from the third type because it includes a group component on top of the multiple variables of interest and is therefore similar to the second type. In this case, the user selected data for *ID*, *Variable*, *Group*, *Variable 2* and/or *Variable 3*. Here, adding a group can serve either to compare the different variables within a group or to display dyadic results. To avoid overcrowded graphs, the comparison

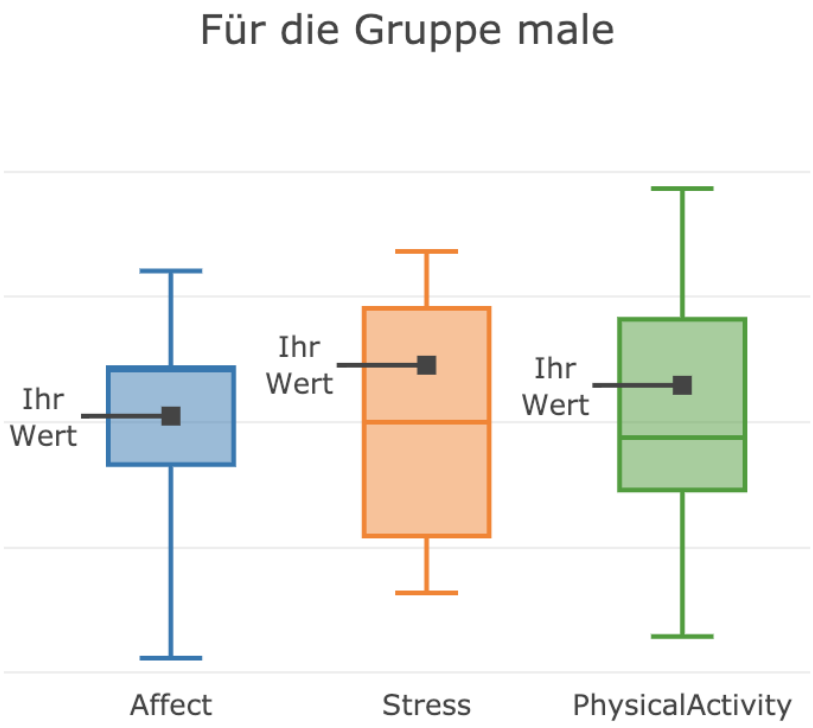


Figure 4.11: Type 4a: Multiple variables within a group

between groups with multiple variables was not considered.

Type 4a: comparing multiple variables within groups

In this case, participants of the group *male* are compared only to other participants in the same group and not to all participants. Again, the values are centered using the z-transformation and the y-axis values is removed (the transformed values do not have an intrinsic meaning). As for the third type, the participant gains four types of information in Figure 4.11. The participant finds out how he does compared to other participants in his group with regards to his *affect*, his level of *stress* and his *physical activity*. For all three items, the participant is in the third quartile. The participant can also see how different is his relative value compared to other males between the three items affect, stress and physical activity.

Type 4b: Dyads and multiple variables

The comparison across different variables is also possible within a dyad. Here, the user provides the ID of the dyad under *ID* and the distinguishable feature under *Group* as well as the main variable in *Variable*, a second variable in *Variable 2* and possibly a third variable in

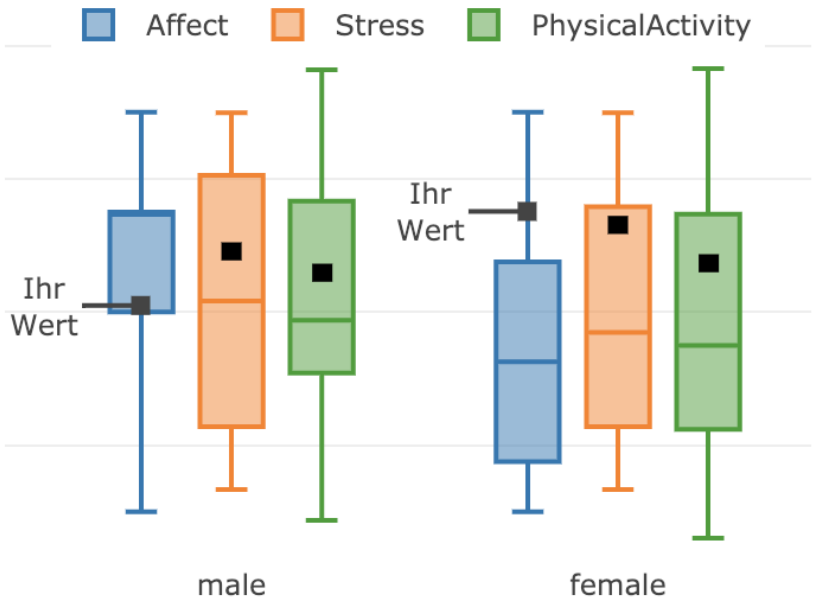


Figure 4.12: Type 4b: Dyads and multiple variables

Variable 3. In Figure 4.12, both the values of the male and the female individuals in the dyad are displayed within the distribution of their own group for each of the three variables *affect*, *stress* and *physical activity*. The participant can see that while both have rather high levels of *stress* and of *physical activity* compared to others in their group, the male participant has a low level of *affect* compared to his group while the female participant is quite high (in the first quartile).

Type 5: time-varying variable

The fifth type handles the case where a time variable is provided with a variable which varies over time. Concretely, the user chooses the *ID*, the *Variable* and *Time*. When a time variable is provided, the message of interest is the fluctuation of the variable. Therefore, the graph displayed is a line graph where the x-axis represent the different time points. A participant's own fluctuation can be compared to the fluctuation of others or to typical values of one's self.

Type 5a: comparison to others

In Figure 4.13, the participant gains information on three elements. First how his or her individual value has varied over the four measurement points. The first time point *Montag* was the highest, folloed be a decrease until the third time point *Mittwoch* and rebound for the last time point *Donnerstag*. The second piece of information is gained

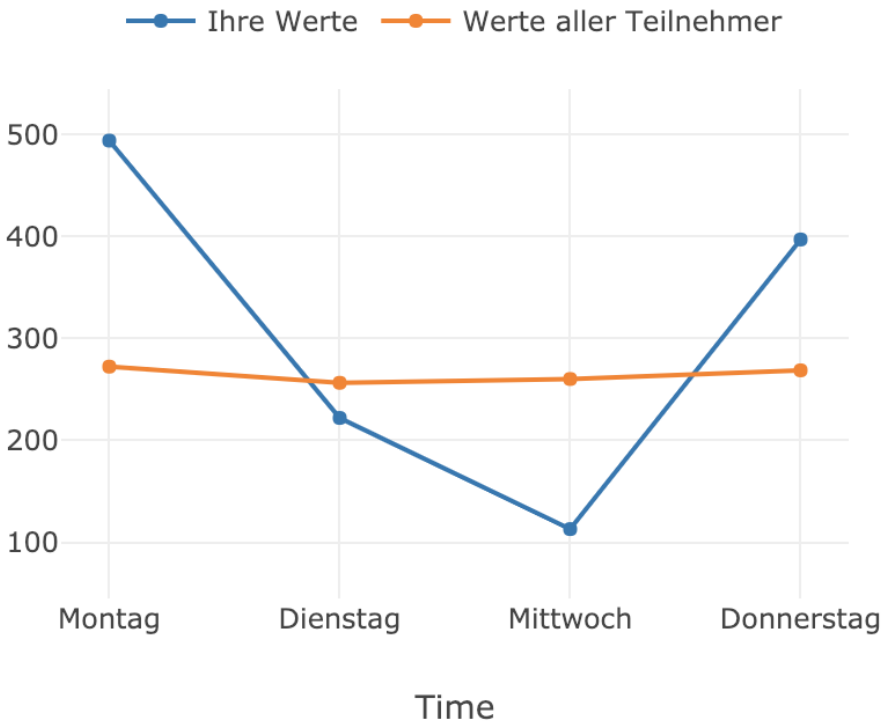


Figure 4.13: Type 5a: own fluctuation compared to others

on the average of all other participants' value at each time point. Figure 4.13 shows that the group mean is quite stable over time so that there are probably no external factors associated with time that would have affected all participants alike. The third piece of information is gained on the comparison for each time point between the own value compared to the mean of all other participants' value. The highest difference to the group mean was on the first time point *Montag* while the participant was closest to the group mean on the second time point *Dienstag*.

Type 5b: comparison to one-self

Comparing an individual fluctuation to the group mean does not always make sense. Sometimes, researchers are interested in within-person comparison, i.e. comparison to oneself. Figure 4.14 informs the participant about three elements. First, the participant sees his or her own fluctuation over time. Second, the participants learns about statistical aggregates of his or her own scores: the full line in the middle is the average of the own values at all time points. The dotted lines are located at the distance of one standard deviation from the mean.

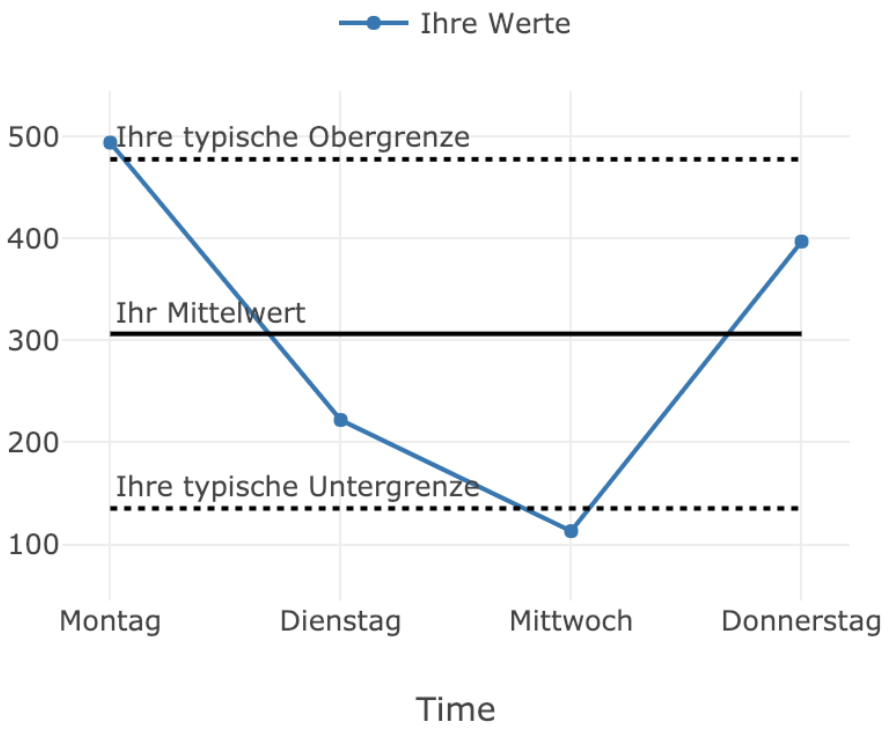


Figure 4.14: Type 5b: own values compared to typical self

To be more easily understood by participants, the are labeled as "own typical higher limit" and "own typical lower limit". The third piece of information concerns the comparison to oneself. In Figure 4.14, the participant was outside of his or her typical range for the time points *Montag* and *Mittwoch* while the two other values are within the range.

Type 6: group differences over time

The sixth type differs from the fifth type by including a group component on top of the variable of interest and the time component. The user selected data for *ID*, *Variable*, *Time* and *Group*. Adding a group can have different purposes which have an impact on how the information needs to be displayed: the group component can relate to a subgroup against which the comparison should be made or to a dyadic structure where the dyad is considered a group of two within which the members are differentiated by a distinguishable feature.

Type 6a: comparing fluctuation to a subgroup

In some cases, it is relevant to compare each participant to his or her subgroup such as *male* and *female* in the example. In Figure 4.15, the participant belongs to the group *male* and gains information about his

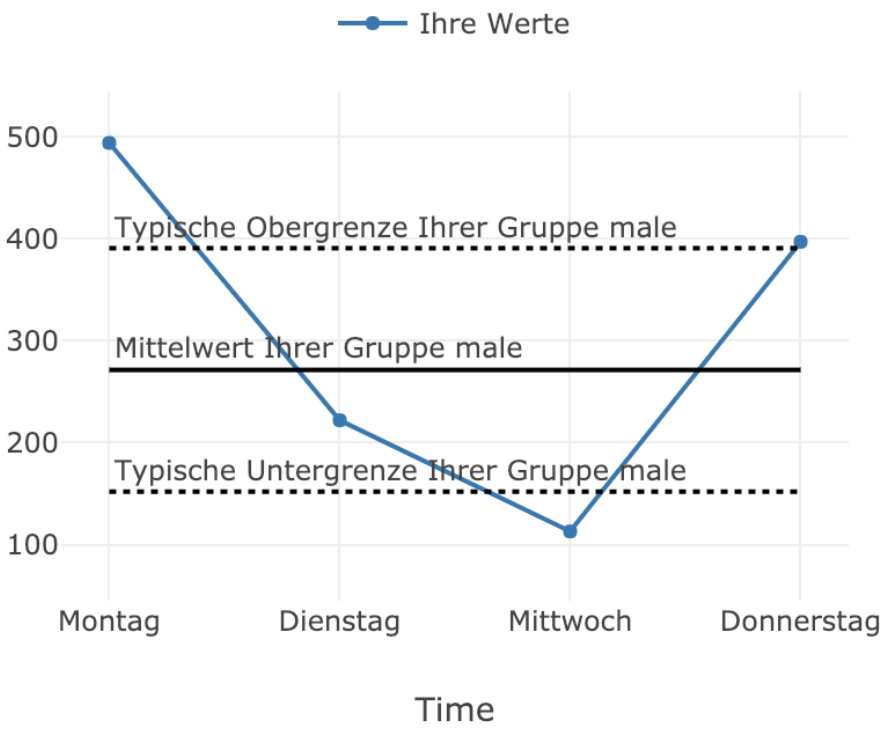


Figure 4.15: Type 6a: Comparing to a subgroup

own fluctuation as well as about how he compares to other participants in his group. The typical range of his group (male participants) is given by the full line and the two dotted lines. The full line represents the average of all male participants over all time points. The dotted lines are located at the distance of one standard deviation from the mean. This range is labeled as the "typical higher limit of your group male" and as the "typical lower limit of your group male".

Type 6b: Fluctuation within the dyad

In the dyadic research, the joint fluctuation of the two members is often of interest. If the user provides the ID of the dyad under *ID* and the distinguishable feature under *Group*, on top of the *Variable* and the *Time*, the fluctuation of both partners is displayed. In Figure 4.16, the dyad is a heterosexual couple and both the values of the male and the female individuals are displayed over time. The participant gains three pieces of information: first his own fluctuation, second the fluctuation of his dyadic counterpart and how the two partners compare at each time point. Figure 4.16 shows that the two partners have opposing values, i.e. when one is high, the other is low and vice-versa.

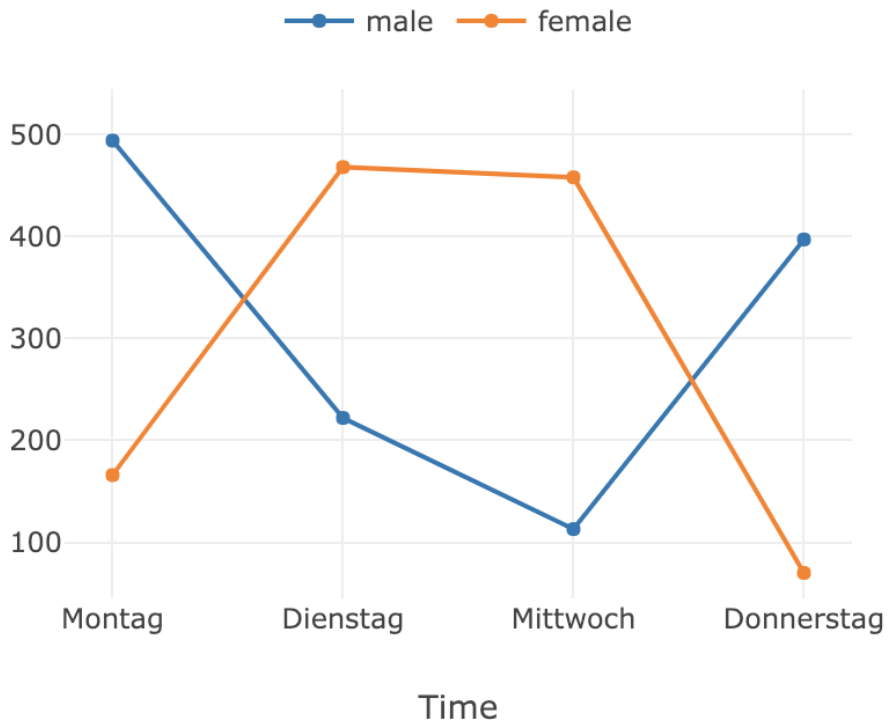


Figure 4.16: Type 6b: Fluctuation within the dyad

Type 7: Multiple variables over time

The seventh type is an extension of the fifth type, where multiple variables are compared with each other overt time. As for the box plots, *Indivi* supports comparison of two or three variables. Since the variables are mostly not measured on the same scale, each variable has an own y-axis with values ranging from the minimum to the maximum of the input values (in the entire data set). Because the focus is not on the individual values but on how the variables fluctuates with each other and because it would make the graph hardly readable, the y-axis are hidden. In Figure 4.17, the participant gains four types of information: first, how *affect* fluctuates over time, second how *stress* fluctuates over time and third how *physical activity* fluctuates over time. Fourth, the participant can see how the three variables compare to each other at each time point. On the first time point *Montag*, all three variables are high. On the second time point *Dienstag*, *stress* and *physical activity* drop while *affect* is still high. On the third time point *Mittwoch*, *stress* goes up while the other two remain low. On the fourth time point *Donnerstag*, *affect* is still low while *stress* and *physical activity* are up again. There seems to be a time-lagged relation between the three variables.

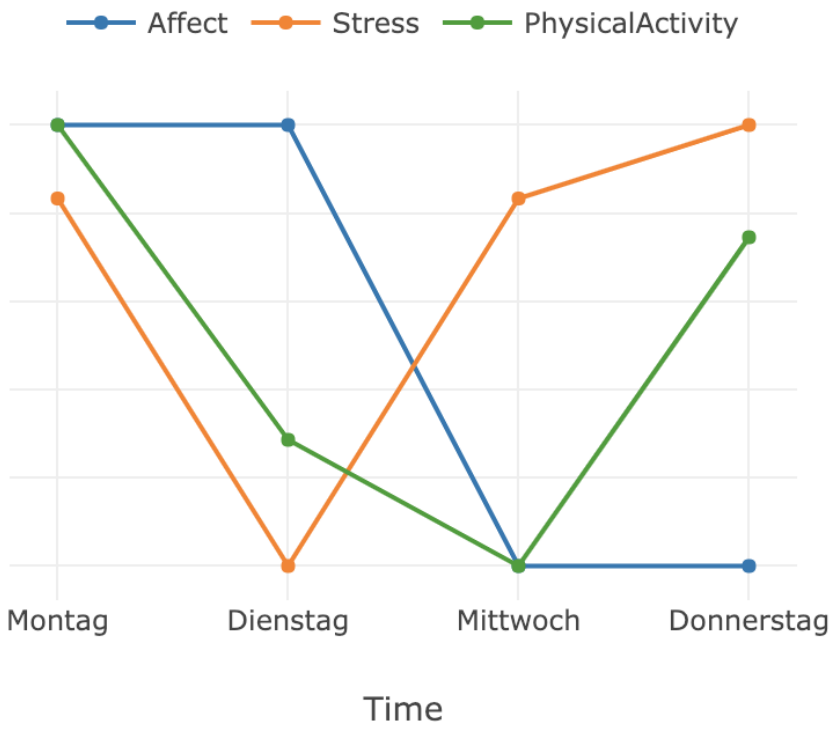


Figure 4.17: Type 7: Fluctuation of multiple variables

Type 8: Group comparison of multiple variables over time

The eighth type combines all possible variable types in one graph. It includes a time component, a group component and multiple variables The user selected data for *ID*, *Variable*, *Group*, *Time*, *Variable 2* and/or *Variable 3*. Compared to the previous graph of type 7, adding a group can serve either to compare the different variables within a group or to display dyadic results.

Type 8a: comparing multiple variables within groups

Participants of the group *male* are compared only to other participants in the same group. To avoid overcrowding the graph, only the average of the group is provided for each variable. The participant gains seven types of information in Figure 4.18. First, the participant sees that his *affect* does not fluctuate much over time, second that his *affect* values are very close to the other males' average, third that his *stress* level fluctuates and is quite high except of the second time point *Dienstag*, fourth that he fluctuates around the group mean but is typically higher than others, fifth that his *physical activity* fluctuates over time, sixth that it fluctuates around the group mean and is as often below as above the mean and seventh that his *affect* fluctuates less that the

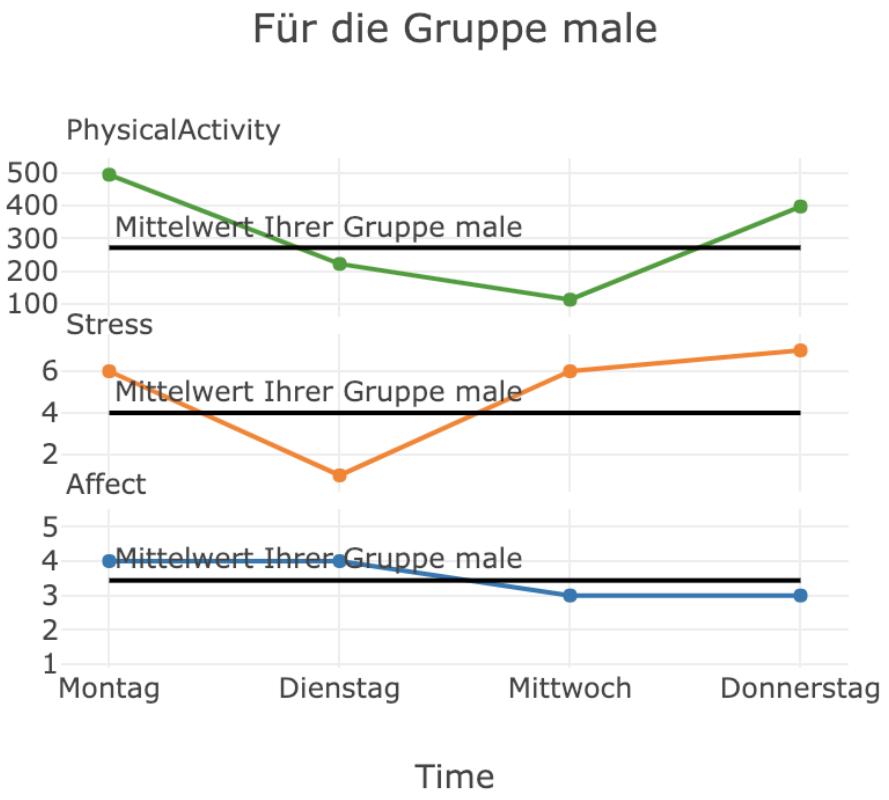


Figure 4.18: Type 8a: Group comparison of the fluctuation of multiple variables

other two which are quite similar in their range of fluctuation.

Type 8b: Dyads and multiple variables

The comparison across different variables over time is also possible within a dyad. Here, the user provides the ID of the dyad under *ID* and the distinguishable feature under *Group* as well as the main variable in *Variable*, a second variable in *Variable 2* and possibly a third variable in *Variable 3*. In Figure 4.19, the fluctuations of both the male and the female individuals in the dyad are displayed for each of the three variables *affect*, *stress* and *physical activity*. The participant gain ten pieces of information: how each of the two partners vary individually for each of the three variables (6 pieces of information) and how they jointly fluctuate for each of the three variables. Figure 4.19 shows that their *affect* is mostly similar and the female participant has typically lower levels, that their *stress* is almost perfectly identical over time and that their levels of *physical activity* is opposite to each other.

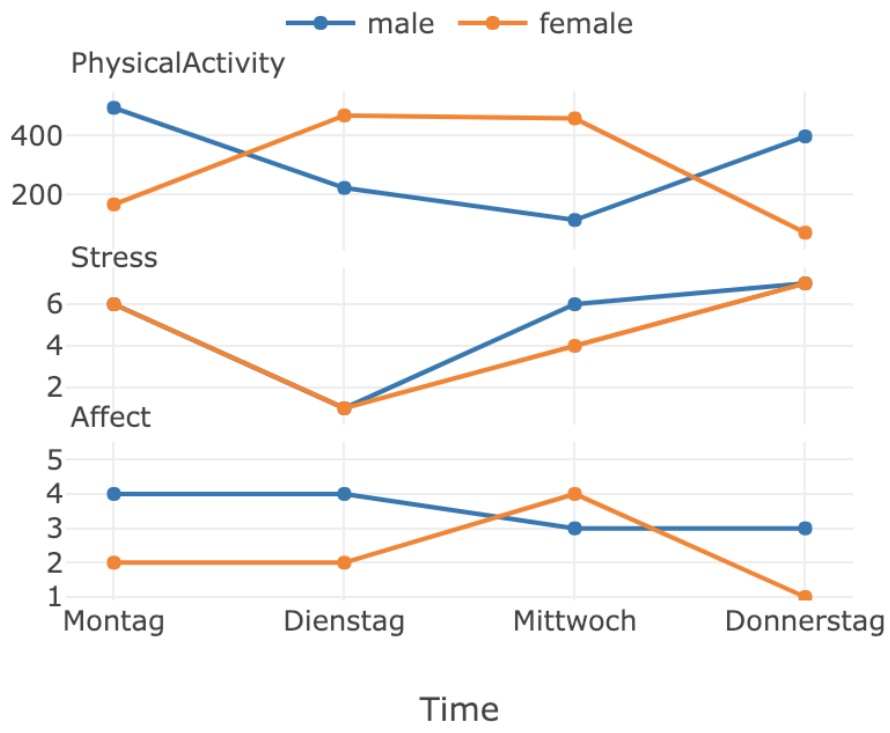


Figure 4.19: Type 8b: Fluctuation of multiple variables in the dyad

The last piece of information gained with this graph is how their joint fluctuation compares between the variables: both partners do not react similarly to their stress levels of changes in affect. More stress is associated with less affect and less physical activity for the woman while it is associated with more physical activity and no change in affect for the man.

A summary of all eight types is referred to in the tree diagram in Figure 3.3. A complete example of a possible individualized feedback in the format of a PDF output can be found in Appendix A. In this example, all possible graphs are displayed and embedded in generic texts that would be individualized for a real feedback.

Chapter 5

Technical implementation

In this chapter, the technical implementation of the *Indivi* tool is presented. The framework as well as the main libraries used in the implementation are discussed and technical choices explained.The code is available online and can be found on GitLab [1] .

JavaScript and React.js as framework for *Indivi*

As part of the prototyping process as well as to test the design principles elicited by this research, a tool enabling to generation of individualized feedback to large number of study participants was created. *Indivi* was developed using JavaScript and React.js[2] as a JavaScript library. React.js provides a framework to simply create single page applications that work well with different media types and screen sizes. With React.js, *Indivi* could easily be made available online as Web-App without the need to download or install a software locally on a computer. React.js also is a community of millions of developers who share tricks, answer questions and contribute to the developments of tens of thousands of packages based on React.js. For such a research project where the focus is not only in engineering the software, relying on a strong community of developers is a notable advantage.

React.js' components

Web-Apps based on React.js are built using components for each core feature. These components can be engineering together to create very complex applications (it is said that Facebook uses more than 30,000 components). The entire application is wrapped in a single mother component that can have children components. Parent and children components communicate via properties (props). Properties are a set

[1] https://gitlab.ifi.uzh.ch/ZPAC/Indivi/
[2] https://reactjs.org

of values including variables and functions. The properties are passed on by the parent component to the child component (top-down) and variables defined in the child component can be passed as arguments of props functions and then be used in the parent component. For communication between sibling components, a communication via the parent component is needed. React.js differentiates between stateful and stateless components. The state of a component is a stable value that hold throughout the component. The state can also be passed on to children components. The state is changed explicitly by calling the `setState` method, telling React.js that the state has changed and triggering a re-render. This allows to actively manage how components work and support the logic behind the component logic of the application.

React.js' life cycle

Using React.js is also made easier by the lifecycle methods underlying the stateful components. The idea behind the lifecycle is that the order of when code is executed within a component can be controlled. This facilitates the communication between parent and children components as well as the asynchronous use of otherwise synchronous functions. The component first mounts, rendering its output for the first time. Directly after the component is mount, first changes can be applied through the `ComponentDidMount` method, which will cause React.js to re-render. After each change in the props, the component will be updated. Here, the methods `shouldComponentUpdate` and `componendWillUpdate` allow to sort what should be updated and to manipulate the new input before rendering. After rendering, the `ComponentDidUpdate` method gives an opportunity to change the state after the update, e.g. comparing a condition before the update with a condition after the update. The update of information within the Web-App is efficient thanks to the use by the React.js framework of a virtual Document Object Model (virtual DOM). After the Web-App has been mounted for the first time, after each change in the Web-App, React.js searches for differences between the current DOM displayed and what the new DOM (after the changes) would look like. React.js then only updates the DOM according to these differences, making it very efficient. This approach is made possible by the use of in-memory cache where the current version of the DOM is stored in a way that can easily be compared with. Nevertheless, this implies a great use of RAM memory.

Session storage

Applications based on React.js also allow for a convenient use of local storage and session storage. Since the need for the *Indivi* tool is very intermittent (specific to the taks of generating feedback) and the

project was set up as a joint venture between different organization structures, the need for and the hosting of a server-side and a database for this application was unclear. The session storage appeared to be a very convenient solution to keep data and user information during a session (as long as the Web-App was open in the browser). With React.js, it is also simple to download the current state of the session storage as well as to update it when a JSON file is provided. Since the researchers had expressed the need to keep a trace of the data and information they had entered during the feedback creation process, using the download and update function of the session storage turned out to be very useful. This allowed to avoid using a server-side, keeping the Web-App as simple as possible and avoiding to have to host a database.

Packages used for *Indivi*

The *Indivi* Web-App was build with the support of three main packages created for or compatible with the React.js framework:

- React-Bootstrap for layout and user-interface elements
- Plotly.js, a d3.js library for rendering and saving plots in javascript
- @react-pdf/renderer, a specific package for rendering PDF documents within the React.js Web-App

Packages: React-bootstrap

React-Bootstrap allow to structure the layout of the Web-App using a flexible grid system composed of columns and rows. It supports the structuring on the rendered output in a way that is simple for the user. The package also includes different graphical user interface elements such as buttons or dropdowns. Panels were used as main graphical element to display text in a box with a title and various color options. Panels offer the option of making the entire box collapsible, which allows for a dynamic rendering of texts depending on the user's current use. A navigation bar was also included to enable the user to see all the process steps at a glance and to jump from one process step to another.

Packages: Plotly.js

As visualization of data is an important part of giving feedback, the Plotly.js library and especially its integration with the React.js framework was used. Two functionalities of the package were primarily used: the rendering of plots within a React.js component and the asynchronous generation of static plots. The former enabled the user to have a glimpse at what the users would see and facilitate the redac-

tion of texts. The latter was used in the process of generating PDF documents and HTML static websites for each participant. The Plotly.js library offers a large variety of plots ranging from scatter plots, line charts, box plots or histograms. This diversity ensured that no technical constraint had to restrain the design choices for visualizing data.

Packages: @react-pdf/renderer

As the output of the feedback generating process has to be some form of document with individualized information for each participant, the generation of PDF documents was a central component. The task was however not to render a PDF document for the user but to generate a document that could be saved and distributed later. The @react-pdf/renderer package is a project with a rather small contributor community which aims at creating PDF documents either on the browser or on the server side. This package was the only one to do enable the generation of documents in the browser at the time of development. Since the application was rather to be a front-end only application, this feature turned out to be essential and the implementation might have been very different without this package. Ensuring the compatibility of this package with future versions of React.js is thus crucial for this implementation.

In order to provide first high-level prototypes and to later finalize a tool that could be used by researchers to generate individualized feedback, the design principles that are the result of this research were implemented in the web-based *Indivi* tool written in JavasSript using the React.js framework that fitted the task particularly thanks to its component-based structure, the efficient rendering of updates using a virtual DOM and the lifecycle approach and the simple handling of session storage. To enhance the Web-App and to perform the core features of the *Indivi* tool, the packages React-Bootstrap, Plolty.js and @react-pdf/renderer were used.

Chapter 6

Testing feedback and discussion

Developing the tool was an iterative process and the tool took several low-level and high-level forms of prototype before reaching its final stage. After reaching this stage, the focus in the interaction with the users switched from understanding the needs and defining requirement towards testing whether the developed tool fulfilled its goal. In this chapter, results from the testing are presented together with the evaluation of the results from the research process will be evaluated.

6.1 Testing process

Goals of the testing

With the testing phase, the objective was to test the fit of the tool in the real-life context, to find hidden bugs and errors, to check whether the flow of the website fits with the practice of providing feedback and to spot confusing elements. The testing took place in two steps. The object of the first testing phase was the tool featuring 50% of the features to make sure that the core of the application was holding its promises. The second testing phase concerned 100% of the tool where all features and functions were implemented.

Participation in the testing

In the first phase, the testing was conducted by three researchers, one senior and two junior researchers, all from the University of Zurich. One tester had not taken part in the interviews. The senior researcher

was the principal. The supervisor of this research and HCI expert also provided extensive feedback on the usability of the tool. In the second phase, the testing was conducted by 9 researchers. All of the researchers that had participated in the interviews also participated in the testing. Two researchers, both juniors, one from Health Psychology and the other from Gerontopsychology, had not participated in the interviews. One had participated in the first testing phase. The supervisor of this research and HCI expert also provided extensive feedback on the usability of the tool. Tester were informed two weeks prior to testing that they would be asked to participate and all confirmed that they would spare some time. Testers were notified again by e-mail when the tool was ready and they were given a deadline of one week.

Response rate

The response quote of the testers after one week was 78%. A reminder was then sent and the response quote raised to 100%. Testers were sent a link to the tool and asked to test the tool with their own data if possible. In case no data were available to the researchers (e.g., data not in the right format), testers were given a data set constructed especially to show all the functionality of the tool. To provide their feedback, a feedback form was attached. Beside stating their name, position and affiliation, the testers were asked what they thought about the workflow of the tool and whether it fitted their usual practice (if any), what they liked and disliked about the tool and what parts were confusing. Finally, they were asked about what future improvements they would like to see.

6.2 Best practice of providing feedback

Finding a common denominator

Since Pope's call [1992] for an integration of feedback as a standard component of Psychological research, little research has been conducted with regards to the effects of feedback and the best practices of providing feedback [Smith et al., 2007, Curry and Hanson, 2010]. The results of the interviews conducted during the need-finding phase confirmed the observation that there is no common denominator or best practice in the current way of giving feedback. After further interviews and a collaborative process centered around co-creation between the interviewer and the interviewee, it was possible to propose a workflow similar to the current practice of different researchers. Once the common denominator was agreed upon by the involved researchers, it was used as the basic structure for *Indivi* where the logic

of the workflow served for the componenets as well as the navigation.

Results from testing

In the testing phase, participants were specifically asked to answer a question about the workflow of the tool. All the testers who had previously provided feedback to participants reported having found the workflow of the tool similar to the way they would manually do it, that all parts needing to be covered were covered and that there were no steps that they thought of as unnecessary. The workflow of the tool was reported to be one of the great features and ensured a familiar and intuitive user experience. Testers who had no experience in giving feedback (junior researchers) reported to have found the tool very intuitive. These junior researchers also reported that the tool had helped them understanding better what was expected from them in giving feedback in future. Testers reported that the tool would save them time and junior researchers highlighted the fact that it would simplify their learning process.

Standardizing to minimize risk

Receiving feedback might not always be an harmless task and providing feedback is often associated with risks such as misunderstanding or misinterpretation [Dixon-Woods et al., 2011]. As avoiding harm is an ethical obligation [American Psychological Association, 2017, Schweizerischen Gesellschaft für Psychologie, 2018], handling risk should be taken seriously into consideration. For sure, the overall presentation is an important factor and might facilitate the reception and provide a positive experience. As suggested by Dixon-Woods and colleagues [2011], using the expertise of a graphic designer ensures a nice design and might have a positive impact on the overall acceptance of the feedback. However, another way of minimizing risk comes from standardizing the procedure. In other industries, for example for health organizations or airlines, the standardization of processes has allowed for less variance, less risks and more safety. Even though the risks at stake are different and might be less critical in the case of receiving feedback, the ethical principles are very strict on the obligation of causing no harm. With a common tool, researchers will naturally provide more similar types of feedback compared as how it is done today. Elaborating *Indivi*, which could become a standard for providing feedback, has given researchers in Psychology a concrete artifact to funnel their knowledge about how feedback should be provided. Hopefully, the development of the feedback practice will continue to evolve. *Indivi* could also serve as platform and concrete object to lead discussions among researchers on how feedback should be provided. Since differences between research fields within Psychology could be overcome,*Indivi* might also be extended to other social sciences who

need to provide feedback to human participants.

6.3 Degree of individualization

Individualizing to minimize risk

The way participants react to feedback and how feedback impact participants differ according to the field of the study, the type of the study, the properties of the participants' population as well as indiviual charactistics such as education, health literary and maybe others [Shalowitz and Miller, 2008]. Beside conducting research on what factors might have an influence on how the participants perceive and interpret feedback, feedback generating tools should therefor include the possibility of providing individualized feedback. Today, it is not only possible on large scales from a technological point of view, it is also wanted by participants [Smith et al., 2007] and ethically right [American Psychological Association, 2017, Lefaivre et al., 2007].

Feedback on *Indivi*

From the need-finding interviews, it became clear that individualized feedback output should take different forms depending on the study and the recipients and that the output should display the own values of the participant both in graphical and textual form. Also, the idea emerged that categories of participants could be created to display entire texts based on the participant's results compared to other participants. As presented in section 4, *Indivi* implements three ways of individualizing the feedback: first displaying individualized graphs where the participant's own value is highlighted, second using placeholders in texts to display the value for each participant while writing the text only once and third categorizing participants according to their scores and writing a specific text for each of the categories. During the testing phase, testers reported to be satisfied with the possibilities given to individualize their output but highlighted that the feedback would still not be as individual as if it was given in person or written specifically for each participant. Concretely, testers reported that the automation of the process radically shrunk the researchers' efforts to provide feedback but that it came at the cost of some individuality.

Degree of individuality

This raises the question of the extent of individuality. From the testers' feedback, it became clear that not every feedback was fully individual. Therefore, some degree of generalization might coexist beside individual result. In this case, individuality is opposed to generalization, when a piece of information does not concern only one individ-

ual but two or more. While the graphs as well as the placeholders in the texts display the individual value of each participant, the interpretation of the graph is not individual but takes place in categories. In *Indivi*, three categories have been implemented. However, following this line of thinking, a fully individual tool would have as many categories than participants. In some way, the degree of individuality could be computed as the number of subcategories K divided by N participants. If K=N, the individuality would be 100%. Reaching 100% seems impracticable, so the degree of individuality seems to be the result of the trade-off between the number of categories, which is important for the participant to receive an interpretation close to his or her own reality, and keeping the workload of researchers low enough for them not to skip it.

6.4 Communicating through graphs

Graph communication

Appropriateness is a key requirement for any feedback and means providing the correct feedback, at the right time and the right way [Lefaivre et al., 2007]. While graph comprehension might be a challenge for participants [Shah and Freedman, 2011], it might not be as difficult for researcher who are used to them. However, communicating through graphs to participants with low graphical literacy requires special skills that cannot be taken for granted for all researchers. Graph communicatiion links three components: choosing the right type of graph, providing domain knowledge and ensuring comprehension independently of the graphical literacy [Shah and Freedman, 2011]. While the latter two can be achieved through text and lay in the expertise of the researchers, choosing the right graph can be challenging in the context of giving written and individualized feedback.

Feedback on graphs

For *Indivi*, two types of graphs were explicitly chosen: the box plot and the line plot. The box plot allows to display the distribution of the data and make a comparison to the group simple enough without loosing relevant information [Weissgerber et al., 2015]. The line plot allows to display within-person information when a time component is present (multiple data points for each participant for the same variable), which would get lost if represented in a box plot. A recurrent feedback from the testing phase concerned the choices of graphs. Testers reported a high adherence to both types of graphs in their respective context (i.e. they could communicate their message through these graphs) but they often wished for more choices of graphs as

users know it from spreadsheet or statistical software. Mostly, the types of graphs wished for included pie charts and bar plots.

Constraining graph choice

While it might sound contradictory to user-centric design, this feedback actually supports the idea that the graph types should be constrained and predefined by *Indivi*. In fact, not all graphs are suited for any type of communication [Few, 2012] and especially pie charts and bar plots turned out not to convey the desired message in a simple yet comprehensive manner. Thus, this testing feedback can be interpreted as a sign that choosing the right type of graph is challenging and the a tool such as *Indivi* should constrain and define the type of graph depending on the data provided to prevent the choice of wrong graphs. So, as receiving feedback as a study participant can be overwhelming, especially the longer the time between the study and the feedback, and because participants are primarily interested in their own scores, a clear focus should be made by the researchers on graph comprehension. Beside the right choice of graph type, which can be constrained by *Indivi*, a particular importance should be give to providing sufficient domain knowledge both on a general level as on the each part of the feedback. The graphical literacy of each participant receiving the feedback, which is an unknown component, can be enhanced by providing correct and understandable explanations on the graph type and proving texts to help interpreting the results. In graphs as well as in texts, the right balance should be found between simplicity and completeness.

Chapter 7

Summary and future work

7.1 Summary and contributions

Researcher-participant relation

Research in the field of Psychology is subject to both changes in the number of participants included in studies and in the their interaction with researchers as well as changes in the type and frequency of data collected during the studies. Participants are expected to provide a greater involvement, to share more personal data and to put more efforts in their participation while the return on participation is still. As a result, participants' involvement suffers, recruitment is difficult and retainment rates drop.

Participants' expectations

Providing feedback was found to be an important way to give back to people who participated in Psychological assessments or studies [Pope, 1992]. From their everyday life, participants are used to receiving individualized feedback about their health, their energy consumption or their spending habits [van Berkel et al., 2017]. Fulfilling this latent expectation is expected to drive up the return on participation and to ease the difficulties to recruit and retain participants. Further, participants expect to be treated ethically. Following the recommendations of the American Psychological Association (APA) [2017], providing feedback is linked to the obligation to avoid causing harm to participants, to give participants all the relevant information before they give consent to participate and to provide an opportunity for obtaining information on the nature, results and conclusions of the research. Both participants' expectations and ethical guidelines advocate for more individualized feedback rather than general results.

Simple feedback generation process

For researchers, preparing and providing feedback is associated with large efforts and long hours. The resources needed rise with the increasing number of participants, especially when feedback is provided individually. An important factor that drives the amount of resources is the lack of standard practice for providing feedback. Many considerations are made with respect to the extent, the content and the format of feedback for individual studies rather than at the institutional, national or international level. The task of providing feedback often falls on junior researcher who lack the experience. With *Indivi*, researchers in Psychology do not only have a tool that automates the process of conceiving, creating and outputting feedback but also that standardizes the typical feedback generation process. Researchers can focus on the research questions they want to give feedback on.

High degree of individualization

Indivi outputs a document for each participant that includes graphs and texts with the participant's own value and entire text blocks corresponding to its own results. From a participant's perspective, this guarantees to fulfill the expectation receiving feedback for their individual values and goes beyond the mere communication of results. Both graphical and textual results are embedded in an interpretation text that provides context information and avoids misunderstandings. From a researcher's point of view, the high degree of individualization is obtained at low costs. Generic texts are written once for all participants with the use of placeholders for which the individual values are automatically replaced in the final output. Graphs are also automatically generated and display the individual value. Using categories based on individual values and percentiles of all values, *Indivi* allows to automate the interpretation texts. By classifying participants according to their either *high*, *medium* or *low* scores and by writing a specific text for each category, different context and tones can be used for different groups. It ensures that the communication is appropriate and gives a way to avoid misconceptions.

Constrained choice of graphs

The need for visualizing data is great because graphs can pontentially be a powerful communication tool. The challenge resides in the "pontentially". There are no graphs that can be used in any situation and many graphs tell another story than the one expected [Few, 2012, Murray, 2017, Munzner, 2014]. This challenge also lives in the context of academic researchers with higher graphical literacy since the expertise needed to provide the right type of graphical communication is often not taught [Shah and Freedman, 2011]. *Indivi* constrains the choice of graph to the combination of data the researcher has chosen for a particular part of the feedback. Following the data tree that resulted

from the contextual inquiry (Figure 3.3), the types of variables were exhaustively listed and a graphical representation was developed for each possible combination. Thus, once the researcher has input the variables needed for the feedback, there is no choice of graph left to be made. While it might sound like a lack of freedom, this constraint actually ensures that the graph displayed is in accordance with the message being communicated.

7.2 Limitations

Broader validation

Indivi as a tool for researchers in Psychology offers a common denominator for feedback practices. It was developed jointly with researchers and based on contextual inquiry. However, further developments of *Indivi* or similar projects would benefit from a more integrative approach of co-creation with the researchers as well as a broader sample of psychologists from different fields to increase heterogeneity. Also, the focus on finding a common ground and to standardize the practice over different sub-fields should be emphasized in the need-finding phase. More integration of researchers with more focus on standardization would lead to more iterations and an output closer not only to how researchers give feedback today but to how they wish to give feedback. The research process should plan the testing with real life data and the participating Psychologists should all bring a data set of their own and have a clear purpose for the development of the tool. For example, the participating researchers could come together in a workshop, try *Indivi* in small groups, discuss their views and give direct, consolidated feedback. Further, the literature found and studied was almost exclusively about therapeutic context, when using psychological assessments in the psychologist-client relationship. The researcher-participant relationship could not be studied explicitly.

Warning for extreme cases

The degree of individualization reached by *Indivi* seems to reach an optimum in the trade-off between resources and number of categories. Having only two categories would not enable to discriminate enough between the participants and the feedback would likely be too general. Adding a fourth category would lead to more effort that would not be compensated by more individual answers. However, special cases are not handled explicitly. Two categories are specific for the lowest and the highest quartile but the individuals with the highest or lowest score might deserve a more specific explanation, or at least read reassuring words. Handling differently participants show-

ing alarming results in the process of giving individualized feedback is of utmost ethical importance [Lefaivre et al., 2007]. In the case of feedback on topics related to physical or psychological health, as well as other psychometric properties, research usually exists in terms of thresholds and other alarming signals for a particular test. A feedback generating tool such as *Indivi* could integrate the possibility for the researchers to manually enter these thresholds for every variable so that the individualized feedback of participants with alarming values would not drown in the mass of "standard" participants, signaling to the researchers to take special action for these participants. This could, for example, lead to a warning before downloading the feedback document or to a special field added for such cases within the document with a link to websites where participants can find help. Under the ethical aspect of causing no harm [American Psychological Association, 2017], this would likely be a needed precaution measure.

More types of data

In the development of the data tree (Figure 3.3) that was taken as basis for the choice of graphs, the focus was made on numerical data (both categorical and ordered, both ordinal and quantitative, cf. [Munzner, 2014]). This type of data covers many aspects of Psychological research such as self-reports (questionnaire data), cognition data, biological data and some sensor data. However, research in Psychology tends to include further types of data such as pictures (e.g., neuropsychological images), audio data (e.g., to measure social interaction) and spatial data (e.g., geographic coordinates to measure mobility). *Indivi* does not handle this type of data although computer support would be needed.

7.3 Future work

This project is quite unique in its interdisciplinary focus on researching how to structure and automate the relation between two target groups: researchers in Psychology on the one hand and study particpant on the other hand. In this project, the researchers in Psychology have been both the subject and the object of the research. This project had a clear focus a particular part of the relationship between researchers and participants, the feedback process, but there are many other areas where this relationship could further be studied and where computer support could be beneficial. Building on the results of this project, three areas of future research are outlined.

Designing the study - feedback for change

In many areas of Psychological research, especially in longitudinal studies, researchers are interested in the behavioral, cognitive, emotional or physical changes of the participants. To make these changes visible, researchers design interventions in their studies to see how participants react to external or internal factors. An interesting area for the study of the researcher-participant relationship could include the development of automated and individual feedback as an intervention tool. How do participants change their behavior if the receive a instant feedback, or a daily feedback or no feedback at all, for example in the context of drinking problems, gambling, physical activity or energy-saving? The use of individualized feedback as a mean of behavioral change should be the object of future projects in the study of researcher-participant relationships.

During the study - data collection

With the increased use of smartphones and other electronic devices in everyday life, the process of collecting data in Psychological studies is also increasingly relying on digital solutions (e.g., for ambulatory assessments studies). Participants are often provided with devices during the time of the study on which they answer self-reports or where sensor data is collected. From a participant's perspective, this indirect interaction over a digital platform represents the largest part of his or her relationship with the researcher. Therefore, the interaction around data collection should be further studied and tools developed which encompass both ends of this relationship. Considering a platform for data collection as inherent part of the researcher-participant interaction, it might make sense for both parties to integrate the data collection platform to a feedback platform such as *Indivi*. This would have further advantages in terms of data cleaning or feedback access for participant. Such an integration should be studied in order to provide a holistic interaction between researchers and participants.

After the feedback - exchange platform

The researcher-participant relationship does not have to end with the last data collection time point. With the work on feedback provided in *Indivi*, the relation is extended until the delivery of the feedback. Future research in this area could include the question of when feedback should be made available to participant. Following the risk of misunderstanding and misconception inherent to feedback [American Psychological Association, 2017], Dixon-Woods and colleagues [2011] suggest to integrate a contact person in form a telephone or e-mail hotline to the feedback. In interviews conducted in this project, researchers suggested the use of a chat bot to automate this part of the relationship. Further research is needed regarding the interaction between researchers and participants around and after the feed-

back. Also, the hypothesis that having a longer-lasting relationship with their participant is expected to have positive impacts on challenges such as participants' recruitment, retainment and involvement should be further studied.

7.4 Conclusion

This research project aimed at designing a tool for researchers in the field of Psychology to create individualized feedback for the participants of their studies. The goal was to guide them and automate the creation of individualized feedback reports. The tool created was called *Indivi*. Three research questions guided this project: (1) find out the design principles for *Indivi* to follow the current practice of giving feedback, (2) find out the design principles for *Indivi* to output individualized feedback, and (3) infer a good visualization for a specific feedback given the type of data provided.

After eliciting the requirements in an agile and iterative manner, the tool *Indivi* was developed and validated through testing by future users. The results of this research respond to the three research questions. (1) A tool for creating individualized feedback has to follow the current practice of providing feedback. However, a consensus about the best practice is lacking. A general process in form of a workflow was elaborated together with researchers in Psychology. (2) The individualization of feedback is the result of a trade-off between the degree of individualization and the resources needed to provide the feedback. An optimal mix can be found by individualizing the feedback on three levels: individualize graphs and display individual values, individualize texts by using placeholders to input individual values, and classify participants into categories for which a specific text is written. (3) The type of the graph can be inferred from a tree combining all types of variables used to give feedback: time component, group component (including dyads) and number of items. Following this tree, the input data and some user choices, a fitting graph is chosen by *Indivi*.

The researchers and future users involved in this project found that *Indivi* represents an improvement compared to current manual practices. At the time of writing, steps had been taken to further develop *Indivi* as an institutional project of the University of Zurich and its Citizen Science initiative.

Appendix A

Full example of feedback output

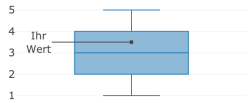
In the following pages, a full example of a potential feedback output from *Indivi* is provided. This example is in PDF-format. Two other formats, booklets and HTML-websites, are also possible as output of *Indivi*. The data used to generate this feedback was generated in order to test the different features of *Indivi*. They do not reflect real data from real participants. The research questions of this example feedback were chosen to show all the possible graphs that *Indivi* supports. The texts are placeholders. As the feedback should be written in German, some parts of the texts are in German while other are in English. If all texts are provided in German in *Indivi*, the output PDF would be entirely in German. This following PDF shows 4 pages on 1 for the convenience of the reader.

Your Feedback

Here is an example of how an individualized feedback could look like.
Here you would first enter some information on the study:
Diese Studie ...

Ihr Affekt

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

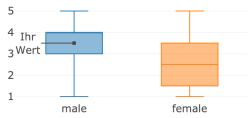


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich in der mittleren Hälfte der gesammelten Antworten. Das bedeutet...

Affekt nach Geschlecht

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

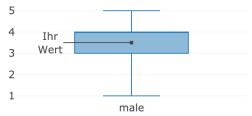


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich in der mittleren Hälfte der gesammelten Antworten. Das bedeutet...

Affekt innerhalb Geschlechtern

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

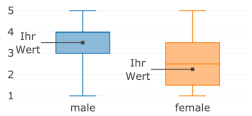


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich in der mittleren Hälfte der gesammelten Antworten. Das bedeutet...

Affekt in der Dyade

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

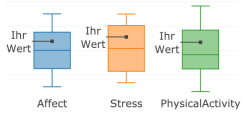


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum.

Personalized text: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

Affekt, Stress und Physische Aktivität

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

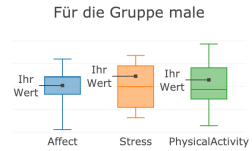


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum. Da die Variablen auf verschiedenen Skalen gemessen wurden, wurden sie transformiert um vergleichbar zu sein (sogennante Z-Transformation).

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich in der mittleren Hälfte der gesammelten Antworten. Das bedeutet...

Affekt, Stress und Physische Aktivität zwischen Geschlechtern

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

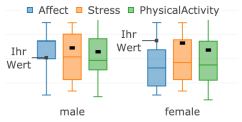


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum. Da die Variablen auf verschiedenen Skalen gemessen wurden, wurden sie transformiert um vergleichbar zu sein (sogennante Z-Transformation).

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich in der mittleren Hälfte der gesammelten Antworten. Das bedeutet...

Affekt, Stress und Physische Aktivität in der Dyade

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

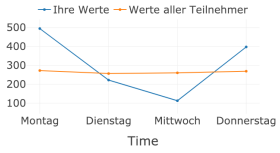


Diese Graphik ist ein Boxplot. Es zeigt die Verteilung der Werte aller Teilnehmer. 50% der Teilnehmer hatten einen Wert innerhalb der Box. 25% der Teilnehmer waren oberhalb und 25% waren unterhalb der Box. Die Linie innerhalb der Box zeigt den Median. Eine Hälfte der Teilnehmer war oberhalb und die andere Hälfte unterhalb. Die Linie oberhalb und unterhalb der Box (Whiskers) zeigen das Minimum und das Maximum. Da die Variablen auf verschiedenen Skalen gemessen wurden, wurden sie transformiert um vergleichbar zu sein (sogennante Z-Transformation).

Personalized text: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

Fluktuation der physischen Aktivität

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

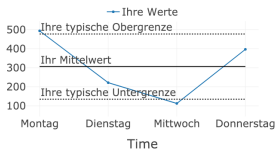


Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich im oberen Viertel der gesammelten Antworten. Das bedeutet...

Eigene Fluktuation der physischen Aktivität

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

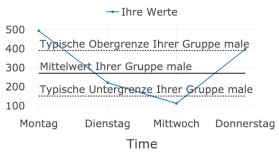


Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich im oberen Viertel der gesammelten Antworten. Das bedeutet...

Fluktuation der physischen Aktivität nach Geschlecht

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

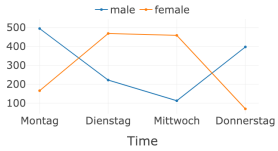


Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich im oberen Viertel der gesammelten Antworten. Das bedeutet...

Fluktuation der physischen Aktivität in der Dyade

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.



Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

Personalized text: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

Affekt, Stress und physische Aktivität über Zeit

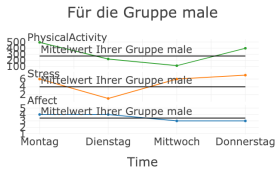
Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich im unteren Viertel der gesammelten Antworten. Das bedeutet...

Vergleich von Affekt, Stress und physische Aktivität in der Gruppe

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

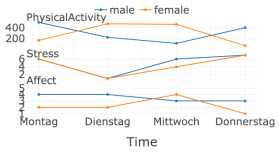


Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

In der Graphik sehen Sie, wo Ihr Wert im Vergleich der anderen Studienteilnehmer liegt. Ihr Wert befindet sich im unteren Viertel der gesammelten Antworten. Das bedeutet...

Affekt, Stress und physische Aktivität in der Dyade

Explanation of the research question: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.



Diese Graphik ist ein Liniendiagramm. Es zeigt die Verteilung der Werte über die Zeit. Die Werte aller Teilnehmer sind der Druchschnitt aller Werte an einem Zeitpunkt.

Personalized text: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim.

At the end, you can provide general conclusions about the study here.
Zusammenfassend, ...

Florian

Bibliography

American Psychological Association. Ethical Principles of Psychologists and Code of Conduct. Technical Report 12, 2017. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.57.12.1060.

David C. Atkins. Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology*, 19(1):98–110, 2005. ISSN 08933200. doi: 10.1037/0893-3200.19.1.98.

Niall Bolger, Angelina Davis, and Eshkol Rafaeli. Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1):579–616, 2003. doi: 10.1146/annurev.psych.54.101601.145030.

Patrick J. Curran and Daniel J. Bauer. The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*, 62:583–619, 2011. ISSN 0066-4308. doi: 10.1146/annurev.psych.093008.100356.

Kyle T. Curry and William E. Hanson. National survey of psychologists' test feedback training, supervision, and practice: A mixed methods study. *Journal of Personality Assessment*, 92(4):327–336, 2010. ISSN 00223891. doi: 10.1080/00223891.2010.482006.

Mary Dixon-Woods, Carolyn Tarrant, Clare J. Jackson, David R. Jones, and Sara Kenyon. Providing the results of research to participants: A mixed-method study of the benefits and challenges of a consultative approach. *Clinical Trials*, 8(3):330–341, 2011. ISSN 17407745. doi: 10.1177/1740774511403514.

Conrad V. Fernandez, Eric Kodish, and Charles Weijer. Informing study participants of research results: an ethical imperative. *IRB; a review of human subjects research*, 25(3):12–19, 2003. ISSN 01937758. doi: 10.2307/3564300.

Stephen Few. *Show Me the Numbers*. Analytics Pres, 2nd editio edition, 2012.

S. T. Fiske, D. T. Gilbert, and G. Lindzey, editors. *Handbook of social psychology*. John Wiley & Sons, vol. 2 edition, 2010.

W Fleeson. Studying Personality Processes: Explaining Change in between-Persons Longitudinal and within-Persons Multilevel Models. In R. W. Robins, R. C. Fraley, and R. F. Krueger, editors, *Handbook of Research Methods in Personality Psychology*, pages 523–542. Guilford Press, New York, 2007.

Amie S. Green, Eshkol Rafaeli, Niall Bolger, Patrick E. Shrout, and Harry T. Reis. Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11(1):87–105, 2006. ISSN 1082989X. doi: 10.1037/1082-989X.11.1.87.

Gabriella M. Harari, Sandrine R. Müller, Min SH Aung, and Peter J. Rentfrow. Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, 18:83–90, 2017. ISSN 23521546. doi: 10.1016/j.cobeha.2017.07.018.

Marlies Houben, Wim Van, Den Noortgate, and Peter Kuppens. The Relation Between Short-Term Emotion Dynamics and Psychological Well-Being: A Meta-Analysis. *Psychological Bulletin*, 141(4):901–930, 2015. doi: 10.1037/a0038822. URL `http://dx.doi.org/10.1037/a0038822901`.

Seungmin Jahng, Phillip K. Wood, and Timothy J. Trull. Analysis of Affective Instability in Ecological Momentary Assessment: Indices Using Successive Difference and Group Comparison via Multilevel Modeling. *Psychological Methods*, 13(4):354–375, 2008. ISSN 1082989X. doi: 10.1037/a0014173.

Rogier A. Kievit, Willem E. Frankenhuis, Lourens J. Waldorp, and Denny Borsboom. Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, 4:1–14, 2013. ISSN 16641078. doi: 10.3389/fpsyg.2013.00513.

Marie Josée Lefaivre, Christine T. Chambers, and Conrad V. Fernandez. Offering parents individualized feedback on the results of psychological testing conducted for research purposes with children: Ethical issues and recommendations. *Journal of Clinical Child and Adolescent Psychology*, 36(2):242–252, 2007. ISSN 15374416. doi: 10.1080/15374410701279636.

M. R. Mehl, J. W. Pennebaker, D. M. Crow, J. Dabbs, and J. H. Price. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments & Computers*, 33(4):517–523, 2001. ISSN 0743-3808. doi: 10.3758/BF03195410.

Tamara Munzner. *Visualization analysis and design*. AK Peters/CRC Press., 2014.

Scott Murray. *Interactive Data Visualization for the Web: An Introduction to Designing with. "* O'Reilly Media, Inc.", 2017. ISBN 9781449339739.

J. B. Nezlek. *Diary methods for social and personality psychology*. Sage Publications, London, 2012.

Kenneth S. Pope. Responsibilities in providing psychological test feedback to clients. *Psychological Assessment*, 4(3):268–271, 1992. ISSN 10403590. doi: 10.1037/1040-3590.4.3.268.

Nilam Ram and Denis Gerstorf. Time-Structured and Net Intraindividual Variability: Tools for Examining the Development of Dynamic Characteristics and Processes. *Psychology and Aging*, 24(4): 778–791, 2009. ISSN 08827974. doi: 10.1037/a0017915.

Schweizerischen Gesellschaft für Psychologie. Ethische Richtlinien für Psychologinnen und Psychologen. Technical report, 2018.

Priti Shah and Eric G. Freedman. Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, 3(3):560–578, 2011. ISSN 17568757. doi: 10.1111/j.1756-8765.2009.01066.x.

David I. Shalowitz and Franklin G. Miller. Communicating the results of clinical research to participants: Attitudes, practices, and future directions. *PLoS Medicine*, 5(5):0714–0720, 2008. ISSN 15491277. doi: 10.1371/journal.pmed.0050091.

S. Shiffman, A. A. Stone, and M. R. Hufford. Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4:1–32, 2008. doi: http://doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Martin J. Sliwinski. Measurement-Burst Designs for Social Health Research. *Social and Personality Psychology Compass*, 2(1):245–261, 2008. ISSN 1751-9004. doi: 10.1111/j.1751-9004.2007.00043.x. URL `http://doi.wiley.com/10.1111/j.1751-9004.2007.00043.x`.

Steven R. Smith, Chauntel M. Wiggins, and Tad T. Gorske. A survey of psychological assessment feedback practices. *Assessment*, 14(3): 310–319, 2007. ISSN 10731911. doi: 10.1177/1073191107302842.

T. J. Trull and Ulrich W. Ebner-Priemer. Ambulatory Assessment. *Annual Review of Clinical Psychology*, 9(2):151–176, 2013. ISSN 10169040. doi: http://doi.org/10.1146/annurev-clinpsy-050212-185510.

Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, 50(6):1–40, 2017. ISSN 03600300. doi: 10.1145/3123988. URL `http://dl.acm.org/citation.cfm?doid=3161158.3123988`.

S. Weintraub, S. S. Dikmen, R. K. Heaton, D. S. Tulsky, P. D. Zelazo, P. J. Bauer, and N. A. Fox. Cognition assessment using the NIH Toolbox. *Neurology*, 80(Issue 11, Supplement 3):S37–S40, 2013. ISSN 0028-3878. doi: 10.1212/WNL.0b013e3182876e0a. URL `http://www.neurology.org/cgi/doi/10.1212/WNL.0b013e3182876e0a`.

Tracey L. Weissgerber, Natasa M. Milic, Stacey J. Winham, and Vesna D. Garovic. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biology*, 13(4):1–10, 2015. ISSN 15457885. doi: 10.1371/journal.pbio.1002128.

Typeset December 31, 2018