8-1-2022

Assessing the Degree of Ecological Validity of Your Study: Introducing the Multidimensional Assessment of Research in Context (MARC) Tool

Sandra Naumann
Humboldt-Universität zu Berlin

Michelle L. Byrne
University of Oregon

Alethia de la Fuente
Universidad de Buenos Aires

Anita Harrewijn
National Institute of Mental Health and Neuro Sciences

Tehila Nugiel
The University of Texas at Austin

See next page for additional authors

Assessing the degree of ecological validity of your study: Introducing the Multidimensional Assessment of Research in Context (MARC) Tool

Sandra Naumann[1]*, Michelle L. Byrne[2, 3]*, Alethia de la Fuente[4-6]*, Anita Harrewijn[7, 8]*, Tehila Nugiel[9]*, Maya Rosen[10]*, Nienke van Atteveldt[11]*, Pawel J. Matusz[12-14]*[†]

1. Berlin School of Mind and Brain & Institute of Psychology, Humboldt-Universität zu Berlin, Germany

2. Department of Psychology, University of Oregon, Eugene, Oregon, USA

3. Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Victoria, Australia

4. Buenos Aires Physics Institute (IFIBA) and Physics Department, University of Buenos Aires, Argentina.

5. Institute of Cognitive and Translational Neuroscience (INCYT), INECO Foundation, Favaloro University, Buenos Aires, Argentina.

6. National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina.

7. Emotion and Development Branch, National Institute of Mental Health, Bethesda, Maryland, USA

8. Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

9. Department of Psychology, The University of Texas at Austin, Austin, Texas, USA

10. Department of Psychology, Harvard University, Cambridge, MA, USA

11. Section of Clinical Developmental Psychology, Research Institute LEARN!, Institute of Brain and Behavior, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam

12. The MEDGIFT Lab, Information Systems Institute, School of Management, University of Applied Sciences & Arts Western Switzerland (HES-SO Valais), TechnoPôle 3, Sierre 3960, Switzerland

13. The LINE (Laboratory for Investigative Neurophysiology), Department of Radiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland

14. Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

* equal contribution; [†] corresponding author: pawel.matusz@gmail.com, +41 586069126

RUNNING HEAD: **Multidimensional Assessment of Research in Context (MARC) tool**

Keywords: ecological, validity, ethological, neuroscience, questionnaire, real-world, naturalistic

Acknowledgments: This work was the result of the 2019 Flux Society Annual Meeting's Pre-Conference workshop "Beyond the lab: Translating developmental neuroscience" organized by Lucia Magis Weinberg, &

Natasha Duell & Jennifer Pfeifer, funded by HopeLab, Bezos Family Foundation & Center for Translational Neuroscience (University of Oregon). This work was supported by the Swiss National Science Foundation grant (no. PZ00P1_174150). We thank Isabel Dziobek for her extensive comments on the original version of this manuscript.

ABSTRACT

In cognitive neurosciences, fundamental principles of mental processes and functional brain organization have been established with highly controlled tasks and testing environments. Recent technical advances allowed the investigation of these functions and their brain mechanisms in naturalistic settings. The diversity in those approaches has been recently (Matusz et al. 2019a) classified via a three-category cycle including *controlled laboratory, partially naturalistic laboratory,* and *naturalistic real-world research*. Based on this model, we developed the Multidimensional Assessment of Research in Context (MARC) tool to easily delineate the approach researchers have taken in their study. MARC provides means to describe the degree of ecological validity for each component of a study (e.g. sample, location, stimuli, measures, etc.) and the study's location on the cycle. The tool comprises seven questions concerning study's characteristics. It outputs a summary of those and a compass plot, which can be used for presentations, pre-registration, grant proposals, and papers. It aims to improve drawing conclusions across studies, and raise awareness about generalizability of research findings.

Lay Abstract

We developed a questionnaire - the Multidimensional Assessment of Research in Context (MARC) tool – to assess ecological validity (the degree in which they capture the real-world behavior they aim to capture) of psychological and neuroscientific studies. MARC enables researchers to explicitly report the level of ecological validity of each aspect of their study by answering seven questions, e.g. about the task they used to the

involvement of non-research experts. The use of MARC should improve transparency, result interpretation, and theory development.

Cognitive neuroscience has taught us a lot about cognitive functions and how these are represented in the brain. Most of these studies, especially the early ones, have used basic computerized tasks and simple stimuli to exert as much experimental control over the studied process as possible. This type of research has provided a lot of detailed information about cognitive processes and their underlying brain mechanisms, but it also has certain drawbacks. One of the main challenges in cognitive neuroscience is the low ecological validity of paradigms for a range of behaviors and cognitive processes naturally occurring typically outside the laboratory-like contexts (Dziobek, 2012; Shamay-Tsoory and Mendelsohn 2019; Matusz et al. 2019a; van Atteveldt et al. 2018), limiting the interpretation of the results with regard to real-life functioning. Our working definition of ecological validity is “a quality of capturing the specific real-world behavioral and environmental factors a study aims to understand” (Dunlosky et al., 2009; Holleman et al., 2020). Low ecological validity could be related to both person-dependent factors (the limited active role of the participants in lab-based paradigms may e.g. interfere with their sense of agency and with the embodiment of their information processing) and situation-dependent factors (artificial, decontextualized environments may not represent real-world interactions) (for more in-depth discussion, see Shamay-Tsoory & Mendelsohn 2019).

Brunswik (1943, 1955) proposed that to achieve generalizability of results from psychological studies, stimuli and tasks should be sampled just like participants are sampled in psychological studies, i.e., in a way that represents the distribution and intercorrelations of ecological variables in the real world. Notably, he defined this quality of a study to reflect the variability present in the outside world as “representative design”. Brunswick was de facto the originator of the term “ecological validity”. However, his definition of ecological validity (Brunswik, 1955) was not only much more precise than the currently existing, more or less explicit, definitions of this construct. It also described a different construct to the one invoked by this term today. Namely ecological validity according to Brunswik is a relationship between a proximal cue delivered to the senses (e.g. stimulation on the retina) and a distal, present in the outside environment, object. Thus, the term describes the potential utility to the organism of a given cue in achieving its behavioral goals. As such, Brunswik's definition of ecological validity is unlike the modern definitions where the term is used interchangeably with "real-world" research.

What Brunswick also recognized, and what has attracted a lot of interest in psychological research in recent years, is that the external environment is uncertain, which has a bearing on some cues being more relevant to one's behavioral goals than others in a given context. The psychological and cognitive neuroscience research has built on this realization. It recognized the importance of information uncertainty as well as the fact that the utility of a given piece of information can be relative. Namely, relevant dimensions of the environment ("signal") are intertwined with the non-relevant ones ("noise"). Recent accounts point out that to guide behavior effectively, the brain needs to continuously actively weigh and re-weigh particular dimensions rather than outright ignore the non-relevant ones (Nastase et al. 2020). Our brains have been shaped to utilize this multidimensionality, which quality may be critical to emulate to understand many behaviors. However, it has not been consistently recognized throughout the history of cognitive research, thus limiting the generalizability of at least some of this work. This has been recently changing. Nastase et al. (2020) has emphasized the added value of researchers identifying explicitly those manipulations that characterize the boundary conditions of behaviors that occur in the environments outside of the laboratory setting. Notably, as Holleman et al. (2020) has recently pointed out, the behavior of interest should always be specified with regard to its context of interest, as opposed to vague terms of "real-life", "ecological validity" or "naturalistic". As an example, instead of studying "real-world social attention", researchers should aim to study attention in a situation of baking a cake, of sharing a meal or of waiting in a waiting room. Only studies within such a well-defined context can shed light on context-specific and context-generic processes governing attention, and other mental processes, in social situations (Holleman et al. 2020). In other words, even a behavior of attending to stimuli on a screen, in a laboratory setting, with limited number and variability of stimuli and task demands, might be characterized by high ecological validity - if one is interested in processes governing attentive behavior in such a setting, those settings will constitute ecologically valid settings.

Whichever processes and contexts are of interest, ecological validity can be assumed to be particularly low in neuroimaging studies, compared to behavioral experimental studies, because of the lengthy, highly controlled tasks and stimuli as well as the artificial and isolated environment, such as an MRI facility, in which the testing takes place (van Atteveldt et al. 2018). Van Atteveldt et al. (2018) describe four approaches to increase the ecological validity of neuroimaging studies. One approach focuses on using more naturalistic tasks and stimuli, such as videos and social interactions. Such stimuli may better reflect some of the behaviors in the outside world, compared to the stimuli in highly controlled lab experiments, as they typically have meaning and vary in the senses they engage from moment to moment. A second approach involves moving the research to environments outside the laboratory by using portable neuroimaging devices, such as EEG, functional near-infrared spectroscopy, or wearable technology. A third approach focuses on combining tightly controlled lab-based neuroimaging measurements with real-life variables and follow-up studies conducted "in the field", for example in the classroom, for those studies where behaviors and cognition inside the classroom are of interest. Lastly, one can improve the ecological validity of their neuroimaging studies by including stakeholders (e.g. teachers and students in the case of studies on learning and education), and doing so at most or all stages of the research (van Atteveldt et al. 2019). All of these approaches help to bring the research closer to understanding information processing and the involved brain mechanisms in everyday environments, and defining relevant research questions. Finally, from a more analytical viewpoint, Nastase et al. (2020) has suggested that ecological validity of investigations into some behaviors can be improved by formalizing hypotheses as explicit models that can offer quantitative predictions of neural activity under the most naturalistic conditions (for those well-defined behaviors) that are possible, and using findings to generate new predictions tested in naturalistic or more controlled contexts.

Recent technical advances, such as increased computational power and better brain mapping tools, have actually provided researchers with the opportunity to more efficiently analyze data from paradigms and experiments in which behaviors occurring outside the laboratory settings are of interest (Bevilacqua et al. 2019; Vanderwal et al. 2019; Rosenblau et al., 2019). Indeed, many recent cognitive neuroscience studies are now starting to use paradigms reflecting and emulating behaviors occurring outside the laboratory (e.g. Föcker et al. 2019; Matusz et al. 2019b; Peelen and Kastner 2014; Vanderwal et al. 2019), as compared to the studies that first pioneered the field, using limited ranges of stimuli and tasks. The former represent different approaches aimed at increasing ecological validity of investigations of behaviors more characteristics of outside laboratory settings: by making use of dynamic stimuli, such as naturalistic movies (e.g. Vanderwal et al. 2019), or audio-visual objects varied in task-relevance, in multi-stimulus displays (Alsius et al. 2011; Cavallina et al. 2018; Matusz et al.2019b; Turoman et al. 2020a, 2020b); by making use of tasks where attention changes dynamically between focused versus divided attention (Föcker et al. 2019) or of real-world scenes (Peelen and Kastner 2014); or even by studying how brain research impacts perceptions of adolescents and their parents (Altikulaç et al. 2019). These recent studies are valuable as they help bridge research on behaviors characteristics of the environments outside the laboratory, like on the high-street or inside the classroom, across more traditional approaches and those involving the study of these behaviors in veridical external environments (Matusz et al. 2019a; Nastase et al. 2020).

Assessing ecological validity

Now that behaviors and cognitive processes are being studied with paradigms that vary in the amount of control over the stimuli as well as over the environment in which they are studied , it would be helpful to explicitly report the levels of this control and 'naturalness' that researchers chose for their study. Explicitly reporting these qualities for stimuli, task, population sampling, etc., in the study takes the burden off the reader, as it is immediately clear what type of research it is and where the study is positioned in the field. In addition, such assessment improves comparing results of different studies, drawing conclusions based on these studies, and identifying gaps for future research. Furthermore, assessing the level of control and the 'naturalness' of their own work encourages researchers to think about them at the design stage of their future studies, and make them generally more aware of the extent to which the experimental setup they have designed offers high levels of ecological validity for the behavior they are interested in studying. These qualities have the potential to increase the ease of interpretation of and so the applicability of cognitive neuroscience studies overall. Therefore, we developed the Multidimensional Assessment of Research in Context (MARC) tool to assess ecological validity of each component part (sample, stimuli, setting, measures, stakeholders etc.) of psychological and neuroscientific studies. This tool can easily be used for communication among researchers and with reviewers, pre-registration, grant proposals, papers, and meta-analyses.

We based the MARC tool on a conceptualization of neuroscientific investigation as a three-category cycle, as proposed by Matusz et al. (2019a). They argue that neurocognitive processes of interest can be studied using three different approaches that complement one another and are each of equal importance: a controlled laboratory research approach, a partially naturalistic laboratory research approach, and a naturalistic real-world research approach. In the controlled laboratory research approach, the process of interest is studied in the lab, where it is isolated by manipulating only a minimum number of factors in a specific experimental design, and all other factors are held constant. This approach provides maximal control over stimuli and environment, which enables the testing hypotheses about behaviors and cognitive processes in specific and highly defined contexts, and doing so with maximized statistical power. In the partially naturalistic laboratory research approach, these process-specific tasks are still administered in lab-like settings but where those settings more resemble everyday situations in which many behaviors may occur. This could be done by selecting different stimuli (e.g., naturalistic movies, multisensory stimuli, or including goal-irrelevant distractors), task conditions (e.g. dynamically changing task difficulty and familiarity, or giving the impression that the participant is being watched by a peer), and/or lab design (e.g., a lab that is set up to look like a classroom, or virtual reality). This approach provides a closer approximation of how the cognitive and/or socio-emotional process of interest might operate in the real world, while maintaining a certain level of control over it and the contexts within which it is gauged. The results of experiments carried out within this partially naturalistic laboratory approach show how well the hypotheses developed within simplified tasks and with simple stimuli hold in contexts more resembling the real-world. In the naturalistic real-world research approach, the process of interest is measured in real-world situations. This approach enables direct testing of the extent to which lab-generated models hold outside traditional laboratory investigations. It likewise allows researchers to uncover new mechanisms supporting cognitive functions in everyday situations or new factors modulating those functions (Matusz et al. 2019a). All three of these categories are of critical importance to the goal of creating more ecologically valid research. Studying a process of interest across all three approaches is important because only together they can provide a more complete understanding of the process of interest and generate hypotheses for its further investigations.

We want to emphasize here that, in line with our working definition, the controlled laboratory research approach can offer "ecologically valid" evidence regarding a behavior of interest, as it is really the nature of that behavior that determines whether a given context is "ecologically valid". Consequently, studies of behaviors in the environments and situations outside the laboratory do not automatically constitute more ecologically valid contexts for investigation of any behaviors, e.g. looking through a person's pictures on social media will not be forcibly instantly studied in more naturalistic fashion if done outside the laboratory. In this context, the MARC tool can be used to assess whether the study represents a more controlled laboratory approach, partially naturalistic laboratory approach, or more naturalistic approach, but without making a judgement about the study's ecological validity.

Including key aspects of a study that can contribute to ecological research (Schmuckler, 2001). As such, MARC consists of a set of seven questions about the design, tasks, stimuli, measures, participant sampling and stakeholder involvement that will be answered by the researcher. The tool provides examples of answers that would reflect one of the three categories of controlled laboratory-based, partially naturalistic, and naturalistic research approaches. Explicit description of key aspects of a study has the potential to improve comparisons between the study in question and other similar research, which in turn should lead to more ecologically valid research (Schmuckler, 2001). We propose that researchers include the results of MARC in their presentations, pre-registration, grant proposals, meta-analyses, and/or as part of the Methods section of their manuscripts, and this way allow others to understand more easily what category their study belongs to.

Why variety in categories of ecological validity matter in adolescent risk-taking research

Choosing the components and tasks that are most appropriate for any given research question requires the researcher to pay special attention to the ecological validity and the construct validity. Different approaches and testing environments are appropriate for different research questions. Consider a construct that has been of great interest in the field of developmental cognitive neuroscience for over two decades: adolescent risk taking (e.g. Casey et al., 1997; Steinberg, 2005, Galvan et al., 2006). While one is ultimately interested in understanding the causes and consequences of adolescent risky behavior in the real world, there are advantages of studying predictors and outcomes related to risk-taking in a controlled laboratory, partially naturalistic, and fully naturalistic manner. One task that is often used in studies of risk-taking is the Balloon Analog Risk Taking (BART) task, which involves having participants inflate a balloon to earn points. But the more they pump the balloon the greater risk they take of the balloon popping and losing all of their points (LeJuez et al., 2002). This task is administered in a controlled laboratory setting and experimenters can manipulate the parameters to make the task more or less risky. This allows a high level of control when considering the implications of the results. However, how much is this controlled task 1) representative of the specific real-world behavior researchers are interested in (ecological validity), and 2) correlated with other predictors or indicators of that same behavior (convergent validity, or more broadly, construct validity)? We can also measure adolescent risk-taking in a partially naturalistic manner by asking participants directly about their real world behavior (for example, through self- or parent-reports about unsafe levels of alcohol consumption or crossing red traffic lights; e.g. Domain Specific Risk-Taking Scale or DOSPERT, Figner & Weber, 2011. We note that although questionnaires may be less naturalistic measures of behavior than actual behavior measured in observational or experimental studies, they are still useful measures of real-world ‘behavior’ (while the limitations characterizing inference from self-report to behavior, such as demands on introspection, should be always kept in mind). This is especially the case in situations where questionnaires or self-reports, as sparse measures of real-life behavior, are utilized to improve the ecological validity of neural measures of real-life behavior by assessing and improving an association between the two (for details, see Section 5.1 in van Atteveldt et al. 2018).

Critically, risky behavior on the BART task has a low- to medium-strength (r=0.243) link to risky real-world behavior related to motor vehicle safety (Vaca et al., 2013), suggesting that this task may also have a high level of construct validity for that particular risk behavior. In other words, the task produces observed behavior in the laboratory that correlates well with observed real-world behavior outside of the lab; therefore, the task appears to measure what it was designed to measure (high construct validity). In this case, tasks that are more naturalistic also have a higher level of construct validity if the construct we are interested in is real-world behavior. We usually want our tasks to correlate well with real-world behavior, to predict and prevent adverse outcomes in real life. Laboratory tasks can be made more naturalistic by adding components such as peer presence. Recent work has explored the impact of peers and parents on adolescent risk-taking behavior and neural responses in the Stoplight task (Chein et al., 2011), or the adapted Yellow Light Game (Op de Macks, et al. 2018; van Hoorn, et al., 2018), in which participants play a simulated driving game and must decide whether to complete the game faster by speeding through yellow lights at the risk of crashing or stopping. Finally, adolescent risk taking behavior can be explored in an even more naturalistic manner outside of the lab through the use of ecological momentary assessment (EMA) which uses mobile devices through which individuals report about their behavior, which could include risk-taking behavior, and emotions in real-time (Kenny et al., 2016). To capitalize on multi-method approaches, some adolescent risk-taking studies have used factor analysis to combine multiple indicators, including several self-report measures and behavioral tasks, of the risk-taking construct (Harden, et al., 2017), with results suggesting the need for further multi-method assessments of psychological constructs.

These examples serve to illustrate that cognitive psychology and/or neuroscience research benefits from diversity in the level of ecological validity (i.e., a spectrum of study design from high control to more naturalistic settings) of tasks and materials. The controlled laboratory research approach (in our example, the BART) provides maximal control over stimuli and environment, which here enabled the testing of specific and highly detailed hypotheses with maximized statistical power. The naturalistic laboratory research approach (in our example, the Stoplight task or the Yellow Light Game) provided a closer approximation of how the cognitive and/or socio-emotional process of interest might operate in the real world, while maintaining a certain level of control over the stimuli and the environment. The naturalistic real-world research approach (in our example, the EMA) enabled direct testing of the extent to which lab-generated models hold outside traditional laboratory investigations. MARC is a tool to allow researchers to consider and clearly justify where their study (or various study components) lies in terms of ecological validity, all the way from the design to reporting stages.

How to use the MARC Tool

You can use MARC at any point in time in the process of conducting your research - before beginning data collection, after collection, when writing a pre-registration, or when submitting a manuscript. The tool can be found online at: https://evaform-git.herokuapp.com/. When you begin, think of responding to the MARC questions as for a single manuscript. With that in mind proceed through each question. Each question is accompanied by sets of examples that fall into the three categories: controlled, partially naturalistic, and naturalistic. Use these examples as a guide to help you categorize each component of your study. Please note that many projects will consist of components that fall on different points along the controlled, partially naturalistic to naturalistic research approach continuum. If you have multiple components for a particular question (e.g. more than one task), simply enter the number of components for that particular section (e.g. using multiple different stimuli, or carrying the study multiple contexts - e.g., the laboratory and in the "field") and answer the question separately for each component. Be sure to accompany your response with a justification/description of each of your components in the space allotted.

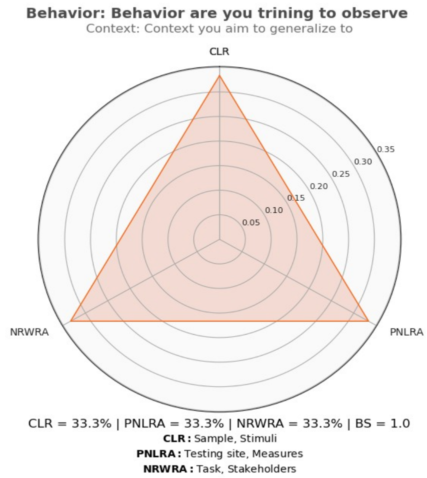

Figure 1. Exemplary compass plot as output of the MARC tool. CLR = Controlled laboratory research; PNLRA = Partially naturalistic laboratory research approach; NRWRA = Naturalistic real-world research approach; BS = Balance Score. The footer of the figure provides information which aspect of the study was evaluated as belonging to which part of the cycle.

The MARC tool allows researchers to answer a series of questions about different aspects of their study (e.g., task, stimuli). When they finish filling out the questionnaire, MARC will produce a compass plot that consists of a "triangle" created by scores along the three axes on a circle (Figure 1). The header of the plot informs about the behavior of interest as well as of the context researchers want to generalize to; the footer includes detailed information as to which aspect of the study contributed to which of the three axes. Each point on the circle represents the extent (here: proportion) to which each of the three categories (controlled, partially naturalistic, naturalistic) that one's study encompasses. For example, a study could be 75% controlled and 25% partially naturalistic or 60% partially naturalistic, 20% naturalistic, and 20% controlled. The orange triangle is drawn by connecting the factor loadings for each category. The balance score (BS) is the normalized area of the triangle, ranging from 0 to 1. Then, a study designed to focus specifically on one factor/category will have a balance score (BS) of 0, while a study designed to equally balance across all three categories will have a BS of 1. We stress that a BS value is not a judgement, i.e., it does not have a 'good' or 'optimum' value. Rather, it represents the balance between the categories of the cycle, providing evidence for the study being more focused on any one category from the 3-category model (Matusz et al., 2019a) versus it being a more balanced mixture of the 3 categories. We refer the reader to the tool itself and to its online version for familiarizing themselves with the attributes that would classify an aspect of a study as belonging to one versus another of the three categories. Crucially, the MARC tool is organized so that your answers to all the questions can be saved as a separate document that you can then attach as part of a pre-registration, or link to in a grant application or a manuscript.

References

Alsius, Agnès & Soto-Faraco, Salvador (2011). Searching for audiovisual correspondence in multiple speaker scenarios. In *Experimental Brain Research, 213*(2-3), 175-183.. DOI:10.1007/s00221-011-2624-0.

Altikulaç, Sibel; Lee, Nikki, C.; van der Veen, Chiel; Benneker, Ilona; Krabbendam, Lydia; & van Atteveldt, Nienke (2019): The Teenage Brain: Public Perceptions of Neurocognitive Development during Adolescence. In *Journal of Cognitive Neuroscience, 31*, 339–359. DOI:10.1162/jocn_a_01332.

Bevilacqua, Dana; Davidesco, Ido; Wan, Lu; Chaloner, Kim; Rowland, Jess; Ding, Mingzhou et al. (2019): Brain-to-Brain Synchrony and Learning Outcomes Vary by Student-Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. In *Journal of Cognitive Neuroscience, 31*, 401–411. DOI:10.1162/jocn_a_01274.

Brunswik, Egon (1955): Representative design and probabilistic theory in a functional psychology. In *Psychological Review, 62*(3), 193. DOI:10.1037/h0047470.

Brunswik, Egon (1943): Organismic achievement and environmental probability. Psychological Review, 50(3), 255. DOI:10.1037/h0060889.

Casey, B. J.; Trainor, R. J.; Orendi, J. L.; Schubert, A. B.; Nystrom, L. E.; Giedd, J. N. et al. (1997): A Developmental Functional MRI Study of Prefrontal Activation during Performance of a Go-No-Go Task. In Journal of cognitive neuroscience 9 (6), pp. 835–847. DOI: 10.1162/jocn.1997.9.6.835.

Cavallina, Clarissa; Puccio, Giovanna; Capurso, Michele; Bremner, Andrew J. ; & Santangelo, Valerio (2018). Cognitive development attenuates audiovisual distraction and promotes the selection of task-relevant perceptual saliency during visual search on complex scenes. In *Cognition*, *180*, 91-98. DOI: 10.1016/j.cognition.2018.07.003.

Chein, Jason; Albert, Dustin; O'Brien, Lia; Uckert, Kaitlyn; Steinberg, Laurence (2011): Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. In *Developmental Science, 14*(2), F1-F10. DOI:10.1111/j.1467-7687.2010.01035.x.

Dunlosky J., Bottiroli S., Hartwig M. (2009). "Sins committed in the name of ecological validity: A call for representative design in education science," in *Handbook of Metacognition in Education*, eds Hacker D. J., Dunlosky J., Graesser A. C. (Abingdon: Routledge;), 442–452.

Dziobek, Isabel (2012): Comment: Towards a more ecologically valid assessment of empathy. In *Emotion Review*, *4*(1), 18-19. DOI:10.1177/1754073911421390.

Eisenberg, Ian W.; Bissett, Patrick G.; Enkavi, Zeynep A. .. & Poldrack, Russell A. (2019): Uncovering the structure of self-regulation through data-driven ontology discovery. In *Nature Communications, 10,* 2319. DOI:10.1038/s41467-019-10301-1.

Figner, Bernd; Weber, Elke U. (2011): Who Takes Risks When and Why? In Curr Dir Psychol Sci 20 (4), pp. 211–216. DOI: 10.1177/0963721411415790.

Föcker, Julia; Mortazavi, Matin; Khoe, Wayne; Hillyard, Steven A.; Bavelier, Daphne (2019): Neural Correlates of Enhanced Visual Attentional Control in Action Video Game Players: An Event-Related Potential Study. In *Journal of Cognitive Neuroscience, 31*(3), 377–389. DOI:10.1162/jocn_a_01230.

Frey, Renato; Pedroni, Andreas; Mata, Rui; Rieskamp, Joerg; & Hertwig, Ralph (2017): Risk preference shares the psychometric structure of major psychological traits. In *Science Advances*, *3*(10), E1701381. DOI:10.1126/sciadv.1701381

Galvan, Adriana; Hare, Todd A.; Parra, Cindy E.; Penn, Jackie; Voss, Henning; Glover, Gary; Casey, B. J. (2006): Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. In The Journal of neuroscience : the official journal of the Society for Neuroscience 26 (25), pp. 6885–6892. DOI: 10.1523/JNEUROSCI.1062-06.2006.

Hackman, Daniel A.; & Farah, Martha J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65-73. DOI:10.1016/j.tics.2008.11.003

Harden, K. Paige; Kretsch, Natalie; Mann, Frank D.; Herzhoff, Kathrin; Tackett, Jennifer L.; Steinberg, Laurence; Tucker-Drob, Elliot M. (2017): Beyond dual systems: A genetically-informed, latent factor model of behavioral and self-report measures related to adolescent risk-taking. In *Developmental Cognitive Neuroscience*, *25,* 221-234. DOI:10.1016/j.dcn.2016.12.007.

Henrich, Joseph; Heine, Steven J.; & Norenzayan, Ara (2010): The weirdest people in the world?. In *Behavioral and Brain Sciences*, *33*(2-3), 61-83. DOI:10.1017/S0140525X0999152X.

Holleman, Gijs A.; Hooge, Ignace T.; Kemner, Chantal & Hessels, Roy S. (2020). The 'Real-World Approach'and Its Problems: A Critique of the Term Ecological Validity. In *Frontiers in Psychology*, *11*, 721. DOI:10.3389/fpsyg.2020.00721.

Holleman, Gijs A., Hooge, Ignace T., Kemner, Chantal & Hessels, Roy S. (2020). The Reality of "Real-Life" Neuroscience: A Commentary on Shamay-Tsoory and Mendelsohn (2019). In *Perspectives on Psychological Science*, 1745691620917354. DOI:10.1177/1745691620917354.

Kenny, Rachel; Dooley, Barbara; Fitzgerald, Amanda (2016): Ecological Momentary Assessment of Adolescent Problems, Coping Efficacy, and Mood States Using a Mobile Phone App: An Exploratory Study. In *JMIR Mental Health*, *3*(4), e51. DOI:10.2196/mental.6361

Lejuez, C. W.; Read, Jennifer, P.; Kahler, Christoper, W.; Richards, Jerry B.; Ramsey, Susan, E.; Stuart, Gregory L.; Strong, David R.; & Brown, Richard A. (2002): Evaluation of a behavioral measure of risk taking:

The Balloon Analogue Risk Task (BART). In *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. DOI:10.1037/1076-898X.8.2.75

Matusz, Pawel J.; Dikker, Suzanne; Huth, Alexander G.; & Perrodin, Catherin (2019a): Are We Ready for Real-world Neuroscience? In *Journal of Cognitive Neuroscience, 31*(3), 327-338. DOI:10.1162/jocn_e_01276.

Matusz, Pawel J.; Turoman, Nora; Tivadar, Ruxandra I.; Retsa, Chrysa; Murray, Micah M. (2019b): Brain and Cognitive Mechanisms of Top-Down Attentional Control in a Multisensory World: Benefits of Electrical Neuroimaging. In *Journal of Cognitive Neuroscience, 31*(3), 412–430. DOI:10.1162/jocn_a_01360.

Nastase, Samuel A.; Goldstein, Ariel; & Hasson, Uri (2020): Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. In *Neuroimage*, In press. DOI:10.31234/osf.io/whn6d.

Op de Macks, Zdena A.; Flannery, Jessica E.; Peake, Shannon J.; Flournoy, John C.; Mobasser, Arian; Alberti, Sarah L.; Fisher, Philip A.; Pfeifer, Jennifer H. (2018): Novel insights from the Yellow Light Game: Safe and risky decisions differentially impact adolescent outcome-related brain function. In *Neuroimage*, *181*, 568-581. DOI:10.1016/j.neuroimage.2018.06.058

Peelen, Marius V.; Kastner, Sabine (2014): Attention in the real world: toward understanding its neural basis. In *Trends in Cognitive Sciences, 18*(5), 242–250. DOI:10.1016/j.tics.2014.02.004.

Rosenblau, Gabriela; O'Connell, Garret; Heekeren, Hauke R.; Dziobek, Isabel (2019): Neurobiological mechanisms of social cognition treatment in high-functioning adults with autism spectrum disorder. In *Psychological Medicine, 50*(14), 2374-2384. DOI:10.1017/S0033291719002472.

Schmuckler, Mark A. (2001): What Is Ecological Validity? A Dimensional Analysis. In Infancy : the official journal of the International Society on Infant Studies 2 (4), pp. 419–436. DOI: 10.1207/S15327078IN0204_02.

Shamay-Tsoory, Simone G.; Mendelsohn, Avi (2019): Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. In *Perspectives on psychological science: a journal of the Association for Psychological Science, 14*(5), 841–859. DOI:10.1177/1745691619856350.

Steinberg, Laurence (2005): Cognitive and affective development in adolescence. In Trends in cognitive sciences 9 (2), pp. 69–74. DOI: 10.1016/j.tics.2004.12.005.

Turoman, Nora; Tivadar, Ruxandra I., Retsa, Chrysa; Maillard, Anne M., Scerif, Gaia, & Matusz, Pawel J. (2020a). The development of attentional control mechanisms in multisensory environments. *BioRxiv*. DOI:10.1101/2020.06.23.166975.

Turoman, Nora; Tivadar, Ruxandra I., Retsa, Chrysa; Maillard, Anne M., Scerif, Gaia, & Matusz, Pawel J. (2020b). Uncovering the mechanisms of real-world attentional control over the course of primary education. *BioRxiv*. DOI:10.1101/2020.10.20.342758

Vaca, Federico E.; Walthall, Jessica M.; Ryan, Sheryl; Moriarty-Daley, Alison; Riera, Antonio; Crowley, Michael J.; & Mayes, Linda C. (2013): Adolescent Balloon Analog Risk Task and Behaviors that Influence Risk of Motor Vehicle Crash Injury. In *Annals of Advances in Automotive Medicine*, *57*, 77–88. PMID:24406948.

van Atteveldt, Nienke; Tijsma, Geertje; Janssen, Tieme; & Kupper, Frank (2019). Responsible Research and Innovation as a Novel Approach to Guide Educational Impact of Mind, Brain, and Education Research. *Mind, Brain, and Education*, *13*(4), 279-287. DOI:10.1111/mbe.12213.

van Atteveldt, Nienke; van Kesteren, Marlieke T. R.; Braams, Barbara; Krabbendam, Lydia (2018): Neuroimaging of learning and development: improving ecological validity. In *Frontline Learning Research, 6*(3), 186–203. DOI:10.14786/flr.v6i3.366.

van Hoorn, Jorien; McCormick, Ethan M.; Rogers, Christina R.; Ivory, Susannah L., Telzer, Eva H. (2018): Differential effects of parent and peer presence on neural correlates of risk taking in adolescence. In *Social Cognitive and Affective Neuroscience, 13*(9), 945–955. DOI:10.1093/scan/nsy071

Vanderwal, Tamara; Eilbott, Jeffrey; Castellanos, F. Xavier (2019): Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. In *Developmental Cognitive Neuroscience, 36*, 100600. DOI:10.1016/j.dcn.2018.10.004.

Appendix 1

Multidimensional Assessment of Research in Context (MARC) Tool

Below you find several short questions regarding your research project. After submitting the form, a summary with your answers and descriptions accompanied by a compass plot will appear. The graph indicates the match of your research idea with three categories that are controlled laboratory-based, partially naturalistic laboratory and naturalistic real-world research. For every category, we added one main example in brackets and some other examples in the tables. The tool can be found online at: https://evaform-git.herokuapp.com/.

1a) What behavior are you trying to observe?

………......

………..

1b) What is the context you aim to generalize to?

………..

………......

2. Thinking about the context you described in 1 a/b, which of the following best describes…
… how your sample reflects that context?

When answering this question, consider how you recruit participants for your study, and whether your sample is representative of, e.g., your region.

[--] Controlled laboratory-based [e.g. Convenience sample, such as undergraduate students at a university]

[--] Partially naturalistic [e.g. Community-based recruitment]

[--] Naturalistic real-world [e.g. A large, nationally representative sample of school districts in a city]

More examples:

Controlled laboratory-based	- Control sample matched only by age and gender - Preclinical studies - Western, Educated, Industrialized, Rich, and Democratic (WEIRD) society - Recruiting children of professors at a university
Partially naturalistic	- Recruiting from one or few local schools - Recruiting an aging sample from several nearby community living facilities
Naturalistic real-world	- Recruiting a large data sample that matches national demographic and socio-economic status characteristic distribution - Large crowd sourced data from public databases (e.g. free narratives, subjective tags, etc.)

3. **Thinking about the context you described in 1 a/b, which of the following best describes…**
… how your testing site reflects that context?

[--] Controlled laboratory-based [e.g. Lab/ clinical testing room]

[--] Partially naturalistic [e.g. Lab set up to look like a classroom]

[--] Naturalistic real-world [e.g. Classroom with little/no experimenter presence and interference into teaching activities]

More examples:

Controlled laboratory-based	- In an M/EEG lab or an MRI scanner/facility - Lab testing room in wet-lab facilities
Partially naturalistic	- More naturalistic stimulation is delivered via VR goggles while wearing M/EEG/fNIRS - In schools and classrooms but not in typical classroom setting (during a normal lesson) - Measuring EEG simultaneously in two participants who are interacting in a lab - Testing ambulatory patients
Naturalistic real-world	- Where the real-world behavior would take place (in the street, market, etc.) - At participant's home - Hyper-scanning during a real concert

4. **Thinking about the context you described in 1 a/b, which of the following best describes…**
   … how your task reflects that context?

Here is a general explanation for what the different categories mean with regard to the task:

- *Controlled laboratory-based: the process of interest is isolated by manipulating only a minimum number of factors in a specific experimental design, and all other factors are held constant*
- *Partially naturalistic: process-specific tasks are used in settings that resemble everyday situations*
- *Naturalistic real-world: the process of interest is measured in real-world situations*

[--] Controlled laboratory-based [e.g. Working memory task for shapes presented on a screen]

[--] Partially naturalistic [e.g. Test of memory after viewing a movie]

[--] Naturalistic real-world [e.g. Memory test of interaction after prolonged delay that involved other activities]

More examples:

Controlled laboratory-based	- Spatial orienting task involving a single target stimulus (a shape) preceded by a single cue/ distractor - Inhibiting a button press to a trained stimulus - Flanker task (responding to the direction of a middle arrow, that is displayed between other arrows pointing in a similar or opposite direction) - Oddball task (responding to a target stimulus in a stream of distractors)
Partially naturalistic	- Selective attention task where both targets and distractors are presented in multi-stimulus arrays (e.g. visual search) and vary across multiple dimensions - Tasks conducted in virtual reality or a room resembling a naturalistic context (e.g. kitchen, a flat or a simple shop) - Social interaction in the lab with a confederate - Clinical neuropsychological task to measure cognitive functions such as fluid Intelligence [Matrix Reasoning, etc], attention [Forward Digit span, etc], executive functions [Trail Making Test Part B, etc], memory [Rey Auditory Verbal Learning Test], social cognition [Reading the Mind in the Eyes Test, etc], etc. - Watching a movie in the MRI scanner - Peer presence during task - giving a speech in the lab in front of confederates/video recording of an audience

Naturalistic real-world	- Selective attention task where both targets and distractors varying across multiple dimensions appear within a veridical external environment (classroom or public place like museum exhibitions) - Observing a child interact with a parent at home - Classroom based and teacher lead curriculum - Free narratives - Observing/transcribing videos of treatment sessions with clinician - Using a fitbit-like motion sensor to measure daily activity - Social network analysis - In-classroom behavior - EMA about social behavior

5. **Thinking about the context you described in 1 a/b, which of the following best describes…**
… how your stimuli reflect that context?

Here is a general explanation for what the different categories mean with regard to the stimuli:
- *Controlled laboratory-based: maximum control over stimuli*
- *Partially naturalistic: some, but not total, control over stimuli*
- *Naturalistic real-world: no control over stimuli*

[--] Controlled laboratory-based [e.g. Static stimuli, typical for perceptual/cognitive studies, like face images]

[--] Partially naturalistic [e.g. Dynamic stimuli, like dynamic faces on video]

[--] Naturalistic real-world [e.g. Fully naturalistically sampled stimuli: people during social interaction]

<u>More examples:</u>

Controlled laboratory-based	-Simplistic stimuli presented multiple times - Stimuli presented one at a time, sequentially - Stimuli varying in their goal-relevance to the performed task (distractor and target stimuli) - Audio clips of phonemes or words - Colored 2D shapes - Pictures of faces with different expressions - Unisensory stimuli
Partially naturalistic	- Rich, naturalistic stimuli with whose properties and their distribution reflect those present in the relevant context - Stimuli varying in their goal-relevance to the performed task (there are distractor and target stimuli) while varying also on other dimensions (see below) - Distractors or target stimulating many senses (visual/auditory, multisensory), - Distractors/targets varying in their familiarity to the observer, being unfamiliar or representing/being connected by a naturalistic object category (animate objects, tools, conspecifics) multisensory - Stimuli whose meaning is dependent on the context (audio or audiovisual clips of social scenarios; watching movies or listening to stories; stimuli presented in VR)
Naturalistic real-world	- Veridical real-world stimuli whose properties and their distribution reflect those present in the relevant context - Disruptions from classmates during a lesson at school - In-place experiences (such as subjective effect of a drug in natural settings, risk activity, etc.)

6. **Thinking about the context you described in 1 a/b, which of the following best describes…**
… how what is being measured reflects that behavior?

[--] Controlled laboratory-based [e.g. Well-understood, well-researched brain correlates of a specific cognitive process, such as the Event-Related Potential (ERP) components P1 or N2 (or a cognitive contrast in fMRI), tested in typical conditions]

[--] Partially naturalistic [e.g. Testing the canonical brain correlates in non-traditional laboratory settings and/or using more portable brain imaging tools, like EEG or fNIRS]

[--] Naturalistic real-world [e. g. Using portable brain imaging tools in veridical external environments to test for canonical brain EEG/ERP 'correlates' of cognitive processes or for spectral features as correlates of mental states (engagement)]

More examples:

Controlled laboratory-based	- Response or accuracy time in rigorous, process-specific lab-based tasks - Environmental variables included only as covariates - Biomarkers such as cortisol level, RNA expression or DNA methylation, etc.
Partially naturalistic	- Using a physical setup inside the testing room or virtual reality - Inter-subject correlational analyses of brain mechanisms during movie watching - In a different / multiple senses - In non-traditional populations (across the lifespan; individuals with non-traditional experience like sensory impairment) - Questionnaire about real-life risk taking - Self / parent report on outside lab behavior as variables of interest
Naturalistic real-world	- Impact of variable of interest for grades or standardized test scores - Topics in free reports extracted by natural language processing (LSA or speech graphs) - Ecological momentary assessment (EMA) ratings of anxiety symptoms - Behavior in classroom - Social network analysis - Real-life behavior data like risk-taking (e.g. alcohol use) or incarceration rates

7**. Are non-research stakeholders involved? (teachers, caretaker, institutions, clinicians)**

[--] Controlled laboratory-based [e.g. Stakeholders only facilitate access to the sample]

[--] Partially naturalistic [e.g. Stakeholders involved in conception OR interpretation/writing up the results]

[--] Naturalistic real-world [e.g. Involvement in conception of project AND interpretation/writing up the results]

More examples:

Controlled laboratory-based	- Practitioners (clinicians, teachers, head teachers, speech therapists) are not involved or involved only through providing the access to the populations of interests
Partially naturalistic	- Practitioners advise on and contribute at some but not all stages of the research project (e.g. result interpretation)
Naturalistic real-world	- Practitioners advise on and contribute to all stages of the research project (e.g. help design, implement, and report on study results)

Is there an intervention component in the study

[--] Yes

[--] No

8. Please indicate where your intervention fits in best.

[--] Controlled laboratory-based [e.g. Children play a game on laptop/ tablet at the lab/ clinic supervised by experimenters and/or parents]

[--] Partially naturalistic [e.g. Children play a game on a laptop/ tablet at home supervised by parents]

[--] Naturalistic real-world [e.g. Children play an online application at home by themselves when they feel like it]

More examples:

Controlled laboratory-based	- Computer paradigm to train participants to look at neutral instead of negative faces
Partially naturalistic	- Training in school with standardized training but outside the regular classroom activities
Naturalistic real-world	- Providing first-line treatment for psychopathology by a trained clinician