Assessing the degree of ecological validity of your study: Introducing the Ecological Validity Assessment (EVA) Tool

Sandra Naumann[1]*, Michelle Byrne[2]*, Alethia de la Fuente[3,4,5]*, Anita Harrewijn[6]*, Tehila Nugiel[7]*, Maya Rosen[8]*, Isabel Dziobek[1*], Nienke van Atteveldt[9]*, Pawel J. Matusz[10,11,12]*†

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

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

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

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

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

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

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

10. Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Sierre, Switzerland

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

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

* equal contribution

[†] Corresponding author: Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO) Valais, Rue Technopole 3, 3960 Sierre, Switzerland, pawel.matusz@gmail.com, +41 276069060

RUNNING HEAD: The Ecological Validity Assessment (EVA) tool

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

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant no. PZ00P1_174150).

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 to investigate those functions and their brain mechanisms in naturalistic settings. The diversity in those approaches has been recently (Matusz et al. 2019a) classified via a three-stage cycle including *controlled laboratory, partially naturalistic laboratory,* and *naturalistic real-world research*. Based on this cycle, we developed the Ecological Validity Assessment (EVA) tool to inform in an easy manner about the approach researchers have taken in their study. It enables objectively describing the study's degree of ecological validity and its location on the cycle. EVA comprises eleven 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 would improve drawing conclusions across studies, and raising awareness for the generalizability of studies.

Lay Abstract

We developed a questionnaire - the Ecological Validity Assessment (EVA) tool – to assess ecological validity (the degree in which their capture the real-world behavior they aim to capture) of psychological and neuroscientific studies. EVA enables researchers to explicitly report the level of ecological validity by answering 11 questions about their study, e.g. about the task they used to the involvement of non-research experts. The use of EVA 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 its drawbacks. One of the main challenges in cognitive neuroscience is the low ecological validity of most paradigms (Dziobek, 2006; 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. We define ecological validity as a quality of representing well the specific real-world behavior the entity aims to represent. 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).

Everyday environments are multidimensional and uncertain, and real-world variables interact nonlinearly with each other. Therein, relevant dimensions of the environment ("signal") are intertwined with the non-relevant ones ("noise"), and, 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, and this realization is not new at all and reappears periodically (Brunswik, 1943; also Gibson 1979; Neisser & Hyman 2000), but its implications have been largely ignored throughout the history of psychological research. Brunswick (1943, 1995), the author of the term "ecological validity", proposed that to achieve generalizability of results from psychological studies, stimuli and tasks should be sampled just like participants are sampled, i.e., in a way that represents the distribution and intercorrelations of ecological variables in the real world. Nastase et al. (2020) added here a few specifications, such as identifying manipulations that characterize the boundary conditions naturally appearing in the real world, formalizing hypotheses as explicit models that can offer quantitative predictions of neural activity under the most naturalistic conditions that are possible, and using findings to generate new predictions tested in naturalistic or more controlled contexts. Notably, as pointed out by Holleman et al. (2019; 2020), the term “ecological validity” (frequently used interchangeably with “real-world research”) should always be well defined. That is, instead of advocating for undefined “higher ecological validity” of psychological research, researchers should specify every time the particular context of cognitive (or social, emotional) behaviors that are relevant to their study and show how these are well represented in that study. 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 more constrained and 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).

Whichever process and context is studied, 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. Stimuli in the outside world, as opposed to stimuli in highly controlled experiments, also typically have meaning and vary in the senses they engage from moment to moment. A second approach involves moving the research to more naturalistic settings 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. 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 (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.

Recent technical advances, such as increased computational power and better brain mapping tools, have provided researchers with the opportunity to more efficiently analyze data from more ecologically valid paradigms (Bevilacqua et al. 2019; Vanderwal et al. 2019; Rosenblau et al., 2019). Indeed, many recent cognitive neuroscience studies are now starting to use more ecologically valid paradigms (e.g. Föcker et al. 2019; Matusz et al. 2019b; Peelen and Kastner 2014; Vanderwal et al. 2019) than the tightly controlled studies that first pioneered the field. These studies represent different approaches aimed at increasing ecological validity: by making use of more ecologically valid stimuli, such as naturalistic movies (e.g. Vanderwal et al. 2019) or audio-visual targets and distractors 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 task conditions that resemble real-world situations, such as focused versus divided attention (Föcker et al. 2019) and 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 more traditional studies and those conducted in highly naturalistic or even veridical external environments (Matusz et al. 2019a; Nastase et al. 2020).

Assessing ecological validity

Now that cognitive processes are being studied with paradigms with different 'levels' of ecological validity, it would be helpful to objectively assess and report the level of ecological validity of a study. Explicitly reporting the level of ecological validity of a 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 ecological validity of their own work might make researchers think about it at the design stage of their future studies, and make them generally more aware of the different levels of ecological validity in cognitive neuroscience research. This has the potential to increase interpretation of and applicability of cognitive neuroscience studies overall. Therefore, we developed the Ecological Validity Assessment (EVA) tool to assess ecological validity 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 EVA tool on a conceptualization of neuroscientific investigation as a three-stage 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: 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 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 of specific and highly detailed hypotheses with maximized statistical power. In the partially naturalistic laboratory research approach, these process-specific tasks are used in settings that resemble everyday situations. 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). Studying a process of interest across all three approaches to ecological validity is important because only together they can provide a more complete understanding of the process of interest and generate hypotheses for its further investigations. In this context, the EVA tool can be used to assess whether a study is an example of the controlled laboratory research approach, the partially naturalistic laboratory research approach, or the naturalistic real-world research approach, which will help relate the study to other studies in the same field.

The EVA tool allows researchers to objectively describe how their study reflects ecological validity of their research project and identify where it fits on the continuum ranging from controlled laboratory research to real-world, fully naturalistic research. EVA consists of a set of eleven questions that will be answered by the researcher about the design, tasks, stimuli, applications, stakeholder involvement and participant sampling. Regarding the latter point, we aim to recognize the need for more participant sampling that matches at least national distribution of demographic and socio-economic status characteristics (Henrich et al. 2010; Hackman & Farah 2009). The tool provides examples of answers that would reflect one of the three categories of controlled laboratory-based, partially naturalistic, and naturalistic research approach. These questions should be answered at the level of an individual paper and we propose that researchers include the results of EVA in their presentations, pre-registration, grant proposals, meta-analyses, and/or as part of the methods section of their manuscripts.

How to use the EVA Tool

You can use EVA at any point in time in the process of conducting your research - before beginning data collection, after collection, when writing a preregistration, or when submitting a manuscript. When you begin, think of answering the EVA questions as answering them 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.

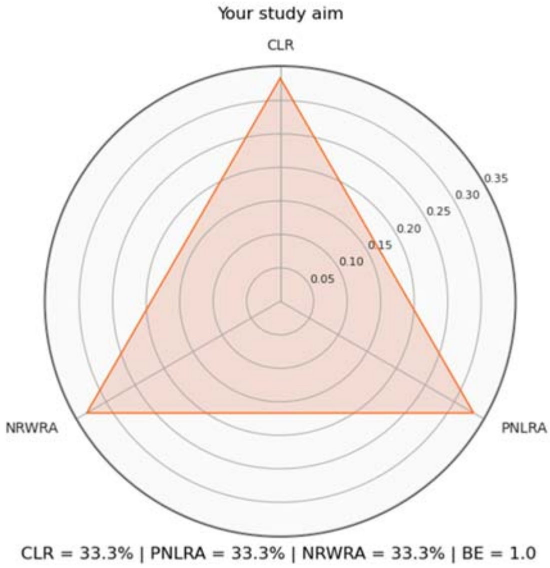

Figure 1. Neurocognitive research can be split into three main components via EVA form as (1) Controlled laboratory research (CLR); (2) Partially naturalistic laboratory research approach (PNLRA); (3) Naturalistic real-world research approach (NRWRA). Expressed as a percentage of the total in the figure. The orange triangle is drawn by connecting the factor loads. The balance score (BE) is the normalized area of the triangle, ranging from 0 to 1. Then, a study designed to focus specifically on one factor will have a balance score (BE) of 0, while a study designed to equally balance factors - a BE of 1. Thus, we stress that a BE value does not have a 'good' or 'optimum' value, but merely aims to summarize and reflect the design characteristics of the study with a single value.

When you finish, EVA will produce a compass plot that consists of a triangle with three points on a circle (Figure 1). Each point on the circle represents the percentage of each of the three categories that your study encompasses. For example, a study could be 75% controlled and 25% partially naturalistic or 60% partially naturalistic, 20% naturalistic, and 20% controlled. We refer the reader to the tool itself and to its online version for familiarizing themselves with the archetypical attributes that would classify an aspect of a study as belonging to one versus another of the three categories. Crucially, EVA 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 linked to in a grant application or a manuscript.

Example: Ecological validity 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. Here we discuss the example of adolescent risk-taking. While we are 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 takes place 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 they 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 alcohol consumption or crossing red traffic lights). We like to note that albeit 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. Moreover, 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 design from high control to more naturalistic) 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. 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. EVA 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.

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.

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.

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

Dziobek, Isabel; Fleck, Stefan; Kalbe, Elke; Rogers, Kimberley; Hassenstab, Jason; Brand, Mathias ... & Convit, Antonio (2006): Introducing MASC: a movie for the assessment of social cognition. In *Journal of Autism and Developmental Disorders*, *36*(5), 623-636. DOI:10.1007/s10803-006-0107-0.

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.

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

Gibson, James J. (1979): *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

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.

Neisser, Ulric; & Hyman, Ira (2000): *Memory observed: Remembering in natural contexts*. Macmillan. New York: W.H. Freeman and Company.

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.

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.

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

Ecological Validity Assessment (EVA) Tool

Below you find some 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.

1. What is the aim of your study? What situation/context do you intend to generalize the results to (e.g. Is your sample representative of your region)?

…….

………......

…….

………......

2. Which methods do you base your main conclusions on? (multiple answers allowed)

- Magneto-/encephalography (M/EEG)
- Electrocardiography (ECG)
- Electrodermal activity (EDA, or GSR)
- functional magnetic resonance (fMRI)
- functional near-infrared spectroscopy (fNIRS)
- Eye-tracking
- Behavioral measures (e.g. RTs)
- ecological momentary assessment (EMA)
- Self-report
- Other:

……..

…….....

3. Describe the kind of stimuli used in your tasks

- 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 in the veridical real world, like persons during a social interaction]

More examples:

Controlled laboratory-based	Partially naturalistic	Naturalistic real-world
-Simplistic stimuli presented multiple times - Stimuli presented one at a time, sequentially - Stimuli varying in their goal-relevance to the performed task (there are distractor and target stimuli) - Audio clips of phonemes or words - Colored 2D shapes - Pictures of faces with different expressions - Unisensory stimuli	- 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	- 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.)

	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)	

4. Describe your experimental approach.

- 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. Test of memory of an interaction after a prolonged delay that involved other activities, including other interactions]

More examples:

Controlled laboratory-based	Partially naturalistic	Naturalistic real-world
- 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)	- 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	- 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

	- Peer presence during task - giving a speech in the lab in front of confederates/video recording of an audience	

5. Where is the testing taking place? (lab, home, classroom, or unconstrained behavior in **a public space, e.g. museum or library)**

- 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. School classroom with little/no experimenter presence and interference into teaching activities]

More examples:

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

6. What are your measures?

- 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	Partially naturalistic	Naturalistic real-world
- 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.	- 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	- 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

Please justify your answer below.

……..

……….......

7. How important your recruitment approach is for generalizing from your results?

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	Partially naturalistic	Naturalistic real-world
- Control sample matched only by age and gender - Preclinical studies - Western, Educated, Industrialized, Rich, and Democratic (WEIRD) society - Recruiting children of	- Recruiting from one or few local schools - Recruiting an aging sample from several nearby community living facilities	- 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

professors at a university		narratives, subjective tags, etc.)

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

Please justify your answer below.

……..

……......

9. Is there an intervention component in the study?

- Yes

- No

If yes, please give some summary information on the intervention, for the reader to understand its fundaments:

What kind of stimuli are you using for the intervention?

……..

……......

How naturalistic is your experimental approach? What is your primary outcome?

……..

………......
Where is the intervention taking place?

……..

………......
To what extent are non-research stakeholders involved?

……..

………......

10. Please indicate where your intervention fits in best.

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

11. Lastly, please indicate which category (controlled laboratory-based, partially naturalistic, and fully naturalistic) you see your research fits best, and state the reasons:

……..

………......