

GUIDELINES FOR PSYCHOEMPIRICAL SOFTWARE ENGINEERING

Daniel Graziotin¹, Xiaofeng Wang¹, Pekka Abrahamsson²

¹ Free University of Bozen-Bolzano, Italy

² Norwegian University of Science and Technology, Norway

SSE 2015, Bergamo, Italy

daniel.graziotin@unibz.it



We study the HUMAN in human aspects of SE



Developers are human beings



Affect – emotions ad moods



Why affect?



Software Development is Intellectual¹ Cognitive²

1. Glass, R. L., Vessey, I., & Conger, S. A. (1992). Software tasks: Intellectual or clerical? *Information & Management*, 23(4), 183–191.
[http://doi.org/10.1016/0378-7206\(92\)90043-F](http://doi.org/10.1016/0378-7206(92)90043-F)
2. Khan, I. A., Brinkman, W., & Hierons, R. M. (2010). Do moods affect programmers' debug performance? *Cognition, Technology & Work*, 13(4), 245–258.
<http://doi.org/10.1007/s10111-010-0164-1>

AFFECT – EMOTIONS, MOODS

Influence^{1,2,3}:

Cognitive activities

Working behaviors

Performance

Software development⁴

1. Ashkanasy, N. M., & Daus, C. S. (2002). Emotion in the workplace: The new challenge for managers. *The Academy of Management Executive*, 16(1), 76–86. <http://doi.org/10.5465/AME.2002.6640191>
2. Fisher, C. D., & Noble, C. S. (2004). A Within-Person Examination of Correlates of Performance and Emotions While Working. *Human Performance*. Lawrence Erlbaum. http://doi.org/10.1207/s15327043hup1702_2
3. Miner, A. G., & Glomb, T. M. (2010). State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes*, 112(1), 43–57. <http://doi.org/10.1016/j.obhdp.2009.11.009>
4. Khan, I. A., Brinkman, W., & Hierons, R. M. (2010). Do moods affect programmers' debug performance? *Cognition, Technology & Work*, 13(4), 245–258. <http://doi.org/10.1007/s10111-010-0164-1>



[Picture Credits](#)

AFFECT – EMOTIONS, MOODS



Interesting,
promising research
avenue for SE.

[Picture Credits](#)

How?

Ψ

How?

4 YEARS OF STUDIES

~~MAIN RESULTS~~



Software engineering lacks in theoretical foundations of affect and guidelines for using psychology

COMMON MISCONCEPTIONS AND ISSUES

Confusing affect with other constructs¹

- Job satisfaction
- Motivation
- Commitment




































1. Graziotin, D., Wang, X., & Abrahamsson, P. (2015). The Affect of Software Developers: Common Misconceptions and Measurements. In *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering* (pp. 123–124). Firenze, Italy: IEEE.
<http://doi.org/10.1109/CHASE.2015.23>

Ooops. Captain, we're lost.

(404. This is not a place where you want to be)



Missing out on validated measurements

SEPTEMBER 2011							
	1	2	3	4	5	6	7
BERND							
MARIKO							
ALEX							
MEIKE							
STEFAN							

Example: Niko-Niko calendar¹

[picture source](#)

1. Sato, D., Bassi, D., Bravo, M., Goldman, A., & Kon, F. (2006). Experiences tracking agile projects: an empirical study. *Journal of the Brazilian Computer Society*, 12(3), 45–64. <http://doi.org/10.1007/BF03194495>



Assuming one truth



“psychologists recognize eight basic emotions, with each positive balanced by a negative”, e.g. “love-hate”¹

“there are six basic emotions or universal emotions: anger, happiness, fear, [..]”²

1. Denning, P. J. (2012). Moods. *Communications of the ACM*, 55(12), 33. <http://doi.org/10.1145/2380656.2380668>

2. Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & García-Crespo, A. (2013). Decisions in software development projects management. An exploratory study. *Behaviour and Information Technology*, 32(11), 1077–1085.
<http://doi.org/10.1080/0144929X.2011.630414>



GUIDELINES FOR PSYCHOEMPIRICAL SOFTWARE ENGINEERING

DEFINING A RESEARCH OBJECTIVE

1

THEORETICALLY FRAMING A RESEARCH

2

SELECTING A VALIDATED MEASUREMENT
INSTRUMENT

3

CONSIDERING PSYCHOMETRIC PROPERTIES

4

ADMINISTERING THE MEASUREMENT
INSTRUMENT CORRECTLY

5

PERFORMING STRONG ANALYSES

6

DEFINING A RESEARCH OBJECTIVE

Scenario

Assessing over a time frame
the **emotional reaction** of a software tool
on developers

Dimensional representation of affect

DEFINING A RESEARCH OBJECTIVE

1

THEORETICALLY FRAMING A RESEARCH

2

SELECTING A VALIDATED MEASUREMENT
INSTRUMENT

3

CONSIDERING PSYCHOMETRIC PROPERTIES

4

ADMINISTERING THE MEASUREMENT
INSTRUMENT CORRECTLY

5

PERFORMING STRONG ANALYSES

6

THEORETICALLY FRAMING A RESEARCH

We are interested in knowing how developers feel in terms of dimensions like pleasure, energy, and dominance.

- We look for a dimensional theory of affect like the one in the PAD models.
 - Pleasure-Arousal-Dominance

DEFINING A RESEARCH OBJECTIVE 1

THEORETICALLY FRAMING A RESEARCH 2

SELECTING A VALIDATED MEASUREMENT
INSTRUMENT 3

CONSIDERING PSYCHOMETRIC PROPERTIES 4

ADMINISTERING THE MEASUREMENT
INSTRUMENT CORRECTLY 5

PERFORMING STRONG ANALYSES 6

SELECTING A VALIDATED MEASUREMENT INSTRUMENT

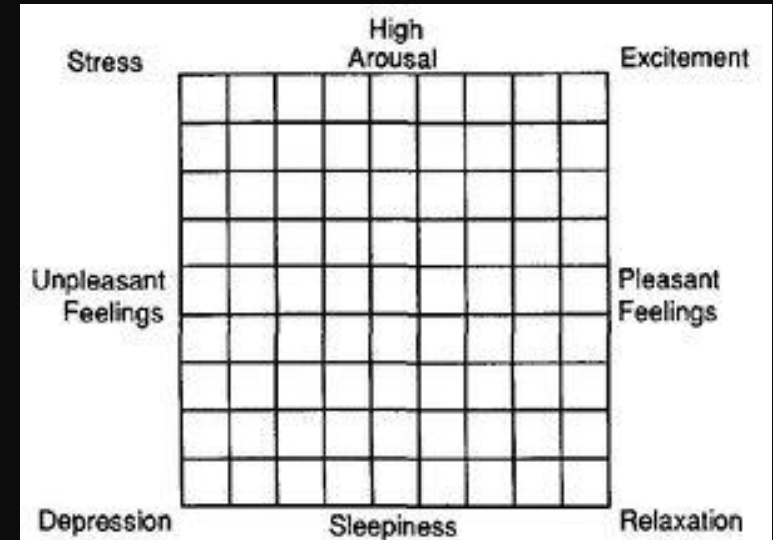
PAD models

Affect Grid¹ notorious

Employed in our field²

Shown to have moderate validity³

Self-Assessment Manikin⁴ (SAM) preferred

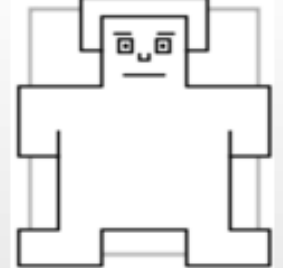
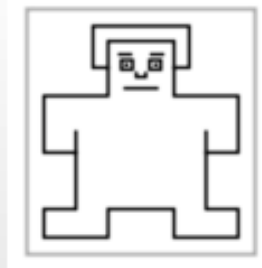
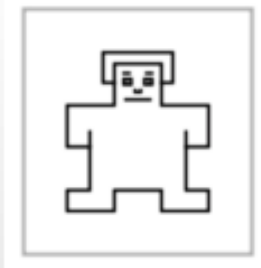
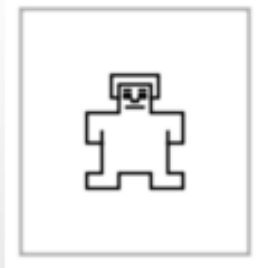
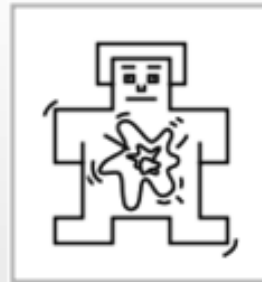
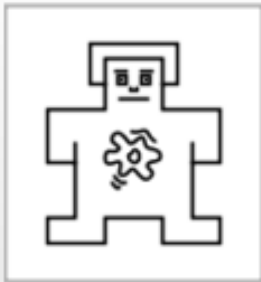
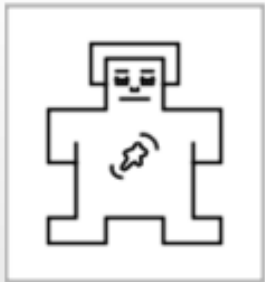
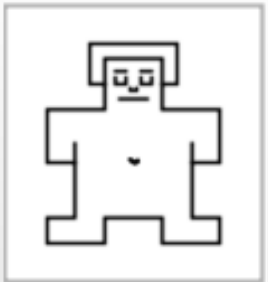
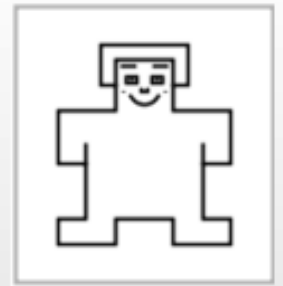
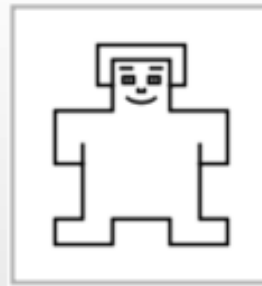
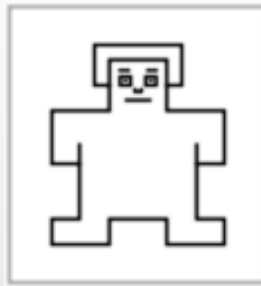
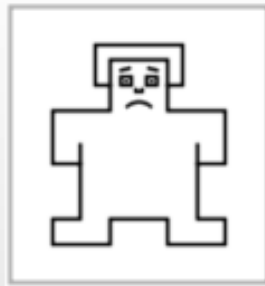
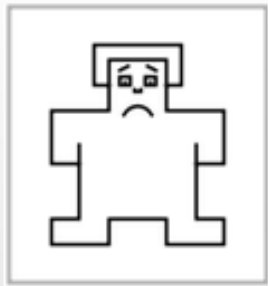


1. Russell, J. a., Weiss, A., & Mendelsohn, G. a. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*. <http://doi.org/10.1037/0022-3514.57.3.493>

2. Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & García-Crespo, A. (2011). Using the affect grid to measure emotions in software requirements engineering. *Journal of Universal Computer Science*, 17(9), 1281–1298.

3. Killgore, W. D. (1998). The Affect Grid: a moderately valid, nonspecific measure of pleasure and arousal. *Psychological Reports*, 83(2), 639–642. <http://doi.org/10.2466/pr0.1998.83.2.639>

4. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [http://doi.org/10.1016/0005-7916\(94\)90063-9](http://doi.org/10.1016/0005-7916(94)90063-9)



Self-Assessment Manikin (SAM)

- 1 DEFINING A RESEARCH OBJECTIVE
- 2 THEORETICALLY FRAMING A RESEARCH
- 3 SELECTING A VALIDATED MEASUREMENT INSTRUMENT
- 4 CONSIDERING PSYCHOMETRIC PROPERTIES
- 5 ADMINISTERING THE MEASUREMENT INSTRUMENT CORRECTLY
- 6 PERFORMING STRONG ANALYSES

CONSIDERING PSYCHOMETRIC PROPERTIES

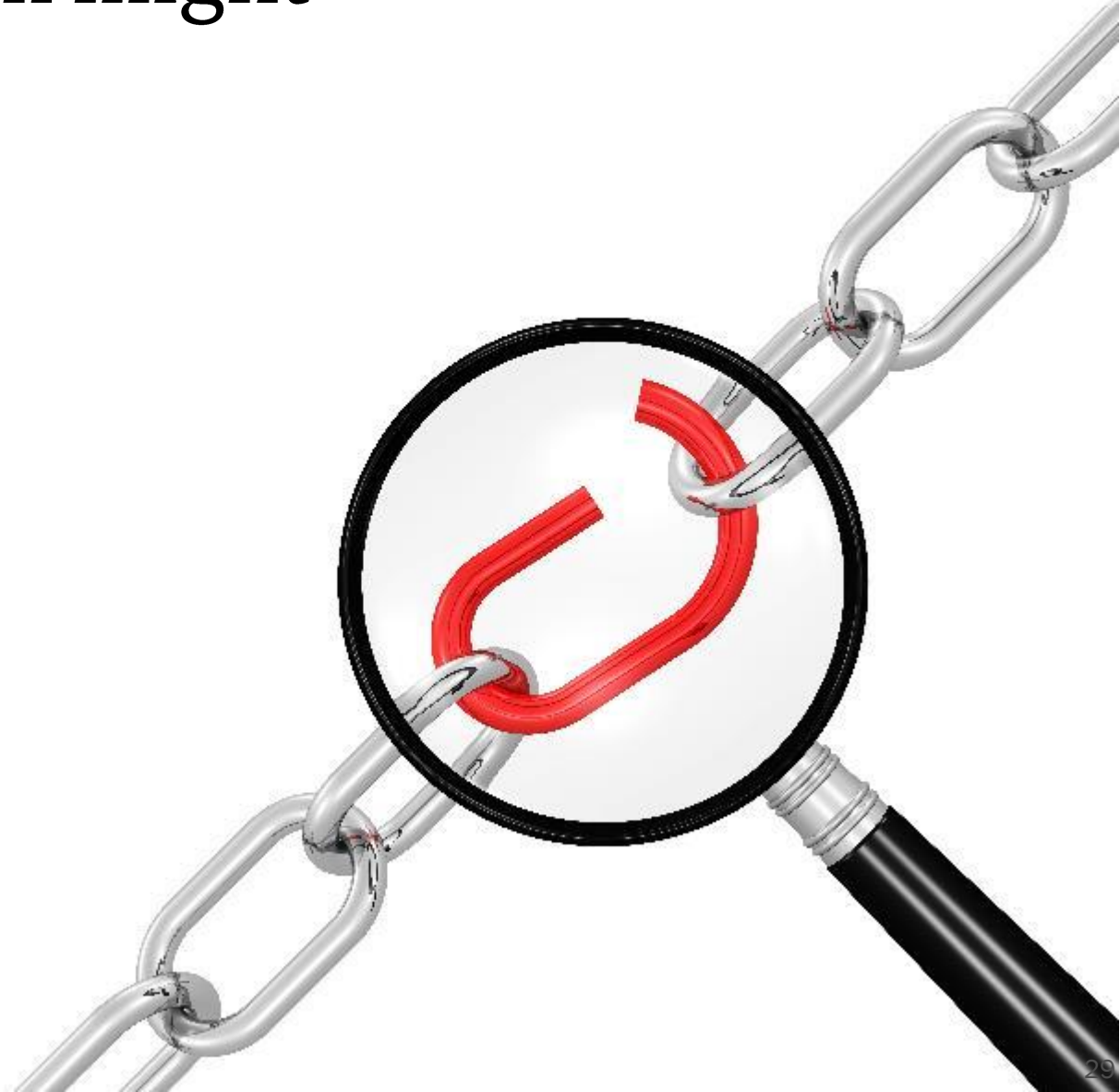
A measurement instrument has to possess acceptable validity and reliability properties

Psychometrics

Design, development, and especially the **validation** of psychological measures.

- Subfield of psychology
 - Not “psychological measurements”

A modification might
require a new
psychometric
assessment



CONSIDERING PSYCHOMETRIC PROPERTIES

"There are six basic emotions [...] Anger, Happiness, Fear, Sadness, Disgust and Surprise [...] We removed 'disgust' and 'surprise' [...] because their translation did not denote [...] everyday emotions."¹

Two issues here

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & García-Crespo, A. (2013). Decisions in software development projects management. An exploratory study. *Behaviour and Information Technology*, 32(11), 1077–1085.
<http://doi.org/10.1080/0144929X.2011.630414>

CONSIDERING PSYCHOMETRIC PROPERTIES

"There are six basic emotions [...] Anger, Happiness, Fear, Sadness, Disgust and Surprise [...]
We removed 'disgust' and 'surprise' [...] because their **translation** did not denote [...] everyday emotions."¹

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & García-Crespo, A. (2013). Decisions in software development projects management. An exploratory study. *Behaviour and Information Technology*, 32(11), 1077–1085.
<http://doi.org/10.1080/0144929X.2011.630414>

CONSIDERING PSYCHOMETRIC PROPERTIES

Scenario

- SAM article provided good psychometric properties¹
- High correlation between SAM items and those of other measurement instruments²
- High reliability across culture and age³
- SAM is one of the most reliable measurement instruments for affective reactions⁴

1. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [http://doi.org/10.1016/0005-7916\(94\)90063-9](http://doi.org/10.1016/0005-7916(94)90063-9)

2. Morris, J. D. (1995). SAM: The Self-Assessment Manikin - An Efficient Cross-Cultural Measurement of Emotional Response. *Journal of Advertising Research*, 35(6), 63–68.

3. Backs, R. W., da Silva, S. P., & Han, K. (2007). A comparison of younger and older adults' self-assessment manikin ratings of affective pictures. *Experimental aging research* (Vol. 31). Taylor & Francis Group.

4. Kim, J., Geason, J. A., Woo, C., & Morris, J. D. (2002). The Power of Affect: Predicting Intention. *Journal of Advertising Research*, 42(3), 7–17.

- 1 DEFINING A RESEARCH OBJECTIVE
- 2 THEORETICALLY FRAMING A RESEARCH
- 3 SELECTING A VALIDATED MEASUREMENT INSTRUMENT
- 4 CONSIDERING PSYCHOMETRIC PROPERTIES
- 5 ADMINISTERING THE MEASUREMENT INSTRUMENT CORRECTLY
- 6 PERFORMING STRONG ANALYSES

Psychometric properties rely on administering the instrument in the same way in each study.

Instructions might influence the participants' responses.

A pair of glasses with dark frames and clear lenses is resting on a white document. A blue pen is also visible, lying horizontally across the bottom right of the document. The document has the words 'INSTRUCTION' and 'MANUAL' printed in large, bold, black capital letters, arranged diagonally from the top left towards the bottom right. The background is a light blue surface.

INSTRUCTION MANUAL

Follow and share

- 1 DEFINING A RESEARCH OBJECTIVE
- 2 THEORETICALLY FRAMING A RESEARCH
- 3 SELECTING A VALIDATED MEASUREMENT INSTRUMENT
- 4 CONSIDERING PSYCHOMETRIC PROPERTIES
- 5 ADMINISTERING THE MEASUREMENT INSTRUMENT CORRECTLY
- 6 PERFORMING STRONG ANALYSES

PERFORMING STRONG ANALYSES

Spend time to understand when you require particular accurate analyses.

Scenario

Repeated measures within-subject with between-subject comparison

- No repeated t-test (or similar)
- Z-score transformation
- No Anova (or RANOVA) -> Linear Mixed Effects models

KEY MESSAGES

- Guidelines are general, and meant to be so
- Respect the reference discipline(s)
- Spend time to read the literature, talk to peers
- Tell reasons behind choices
- **Be transparent**

Thank you for your attention

psych.se

daniel.graziotin@unibz.it