

# Online Supplement

## Appendix A

*Table 1: Search Terms for all relevant aspects of our search query with pseudonyms; asterisks indicate wildcards, quotation marks indicate all relevant permutations of a search term, if wildcards were no option due to an expected increase in noise*

Dimensions of Search Query	Main Search Term	Pseudonyms
(a) Digitalisation	digital*	artificial, computer*, "pc", mobile, iphone*, "app", "apps", web, tablet*, ipad*, media, facebook, twitter, twitch, streaming, secondlife, minecraft, "mmo", "mmorpg", youtube, iTunes, electronic*, game*, gaming, play*, network*, blog*, "3D", cyber, virtual, "VR", "AR", "augmented reality", "virtual reality", interactiv*, internet
(b) AC Subfacet	Culture	cultur*
	Visual Arts	"art", "arts"
	Museum	museum*
	Music	music*, "sing", "singing", melod*, song*, orchestr*, choir*
	Performing Arts	danc*, theatre*
	Literature	"read", "reading"
	Photography	photograph*
	Movies	movie*
	Video Games	game*
	Education	education*
(c) Education	education*	learn*, school*, student*, pupil*, college*, universit*, teach, teachers*, bildung, pedagogic*, social*, enjoyment*, engagement*, immersion*, cognitive, affective, motivation*, behaviour*, behavior*, emotion*, fun*
(d) Quantitative	empirical*, quantitative*	Mean, regression*, correlation*, cluster, factor, path, model, cfa, efa, sem, bayes*, anova, "sample size", experiment*, longitudinal*, latent*, lisrel, mplus, spss, "sas" Stata, "R", cronbach

## Appendix B

Table 2: Reduction of word count (WC) and unique word count (UWC) separately reported for each subsequent cleaning process.

Cleaning processes								
Title	Variable	Corpus before cleaning	Deletion of copyright and non-English titles	Substitution of specific words	Deletion of stop words	Stemming of all words	Deletion of words with 4 hits or less	Total reduction in %
	WC	796,051	739,187	736,817	480,773	470,282	429,828	46.01
	UWC	49,790	36,775	36,768	36,413	26,185	5,320	89.32
	WC/UWC	15.99	20.1	20.03	13.20	17.96	80.79	
	M(WC)	14.33	13.31	13.26	8.66	8.47	7.75	45.92
	SD(WC)	6.49	4.58	4.57	2.93	2.88	2.81	56.70
Abstract	Variable	Corpus before cleaning	Deletion of copyright and non-English titles	Substitution of specific words	Deletion of stop words	Stemming of all words	Deletion of words with 4 hits or less	Total reduction in %
	WC	11,657,752	11,350,979	11,337,303	6,372,003	6,154,580	5,429,095	53.43
	UWC	128,958	123,033	122,997	122,552	92,615	12,409	90.38
	WC/UWC	90.39	92.26	92.18	51.99	66.45	437.51	
	M(WC)	209.85	204.33	204.08	114.72	110.81	97.74	53.42
	SD(WC)	85.17	85.43	85.39	47.67	46.37	40.09	52.93
Keywords	Variable	Corpus before cleaning	Deletion of copyright and non-English titles	Substitution of specific words	Deletion of stop words	Stemming of all words	Deletion of words with 4 hits or less	Total reduction in %
	WC	500,089	--	496,250	455,361	449,025	416,128	16.79
	UWC	31,623	--	31,622	31,359	23,792	5,156	83.70
	WC/UWC	15.81	--	15.69	14.52	18.87	80.71	
	M(WC)	9.00	--	8.93	9.44	9.31	8.63	4.11
	SD(WC)	5.23	--	5.21	4.00	3.96	3.67	29.83
Source Title	Variable	Corpus before cleaning	Deletion of copyright and non-English titles	Substitution of specific words	Deletion of stop words	Stemming of all words	Deletion of words with 4 hits or less	Total reduction in %
	WC	219,523	--	218,924	144,213	138,458	127,525	41.91
	UWC	3,757	--	3,759	3,671	3,214	2,466	34.36
	WC/UWC	58.43	--	58.24	39.28	43.08	51.71	
	M(WC)	3.95	--	3.94	2.60	2.49	2.38	39.75
	SD(WC)	1.85	--	1.85	1.11	1.09	1.03	44.32

## Appendix C

### *Refinement via WC*

Words such as “child”, “children”, “adolescent” or “knowledge” were extracted as significant words for the score of education. Within the facets of aesthetics, for example, we extracted the word “library” for literature as a subfacet of AC activity, and regarding the negative score, we identified words such as “amino”, “acid” and “physics”.

### *Refinement via Term Frequency (tf)*

We also analysed the mean rank of term frequencies across all text objects to refine our list of significant words. This especially revealed significant words linked to digitalisation. For this facet, we extracted words such as “selfie”, “meme”, “troll”, and “gamify”. Within the facet of AC, “Shakespeare” was added as a significant word for performing arts and literature. “Lithium” was extracted as an additional negative significant word.

### *Refinement via tf-idf*

The analysis of words ranked via *tf-idf* resulted only in the extraction of negative significant words by including words such as “antibiotic”, “microbiology”, “graphene”, “fungal”, “orexin” and “midwives”.

### *Refinements after each Iteration*

After the first iteration, we manually screened the top  $n = 50$  words of included and excluded papers for each text object with log-ratios for *tf* and *tf-idf* (cf. Text Mining Statistics) resulted in further positive and negative significant words. On the one hand, “edutainment”, “postmodern”, “avatar” or “multiplayer” were words predominantly occurring in publications that were manually judged to be included. On the other hand, “neural”, “fluency”, “aphasia”, “brain”, “chemistry” or “psychometric” were words mainly

encountered in publications that were manually judged to be excluded. Subsequent iterations of this refinement process resulted in the extraction of additional significant words such as “warcraft”, “gbl”, “adventure”, “choreography”, “drinking”, “kinect”, “exergame”, “kindergarten”, “psychopathology”, “ancient”, “quantum” or “diaspora”. After each refinement, we reiterated the significance scoring process for each text object of each publication.

## Appendix D

*Table 3: Mean publication averaged significant scores in percent for included and excluded publications during manual screening.*

Significance Score	Mean Excluded in %	Mean Included in %
Culture	0.67	1.94*
Visual Arts	0.00	0.14*
Museums	0.03	0.41*
Music	0.17	0.63*
Performing Arts	0.16	0.15
Literature	0.49	1.24*
Film	0.11	0.35*
Photography	0.11	0.57*
Video Games	0.38	5.44*
Digital	4.42	24.22*
AC	2.23	10.96*
Education	9.32	14.01*
Quantitative	1.63	1.67
Positive	17.78	50.93*
Negative	2.34	0.48*

Note: \*: significant difference between included and excluded mean significance scores analysed via t-tests.

## Appendix E

*Table 4: Total and relative frequencies of included publications with significance scores larger 0. All publications including publications research video games on the left side; all publications excluding publications researching video games on the right side.*

Significance Score	Including video games		Excluding video games	
	publications with significance scores > 0	% of publications with significance scores > 0	publications with significance scores > 0	% of publications with significance scores > 0
Culture	332	19.9	249	30.0
Visual Arts	43	2.6	38	4.6
Museums	74	4.4	65	7.8
Music	101	6.1	91	11.0
Performing Arts	48	2.9	36	4.3
Literature	203	12.2	164	19.7
Film	110	6.6	99	11.9
Photography	114	6.8	104	12.5
Video Games	836	50.1	n.a.	n.a.
Digital	1653	99.2	820	98.7
AC	1457	87.5	623	75.0
Education	1501	90.1	712	85.7
Quantitative	1006	60.4	460	55.4
Positive	1666	100.0	831	100.0
Negative	141	8.5	33	4.0

Note: Differences in sum scores and sums of D-ACE facets of subfacets of AC result from documents' significance scores larger 0 in multiple AC subfacets.

## Appendix F

Table 5: Top 10 ranked words by word-topic probability  $\theta$  of each of  $k = 4$  extracted topics of  $n = 831$  documents researching D-ACE excluding documents researching video games.

Rank	Topics			
	1	2	3	4
1	User	Digital	Learn	Culture
2	Online	Read	Computer	Media
3	Social	Video	Education	Internet
4	Socialmedia <sup>a</sup>	Music	Student	Selfie
5	Facebook	Book	Interact	Mobile
6	Behavior	Student	Art	Image
7	Engage	Technology	Teach	Visual
8	Socialnetwork <sup>a</sup>	Children	Museum	Draw
9	Motivation	Write	Virtual	Photography
10	Blog	Cognitive	Technology	Phone
Mean Document-Topic Probabilities	25.98%	24.24%	25.20%	24.58%

Note: <sup>a</sup> : Social media and social network are fixed terms, that have been grouped together via parsing.

Table 6: Top 10 ranked words by word-topic probability  $\theta$  of each of  $k = 6$  extracted topics of  $n = 831$  documents researching D-ACE excluding documents researching video games.

Rank	Topics					
	1	2	3	4	5	6
1	Online	User	Read	Video	Learn	Digital
2	Social	Socialmedia <sup>a</sup>	Student	Music	Education	Culture
3	Internet	Twitter	Book	Computer	Art	Selfie
4	Student	Engage	Interact	Mobile	Virtual	Photography
5	Facebook	Technology	Write	Technology	Teach	Image
6	Behavior	Motivation	Technology	Cognitive	Museum	Media
7	Engage	Visual	Blog	Phone	Interact	Motivation
8	Socialnetwork <sup>a</sup>	Knowledge	Children	Visual	Children	Draw
9	Adolescent	Affect	School	Knowledge	School	Picture
10	Blog	Emotion	University	Youtube	Exhibition	Film
Mean Document-Topic Probabilities	16.75%	16.71%	16.64%	16.45%	16.79	16.66

Note: <sup>a</sup> : Social media and social network are fixed terms, that have been grouped together via parsing.

Table 7: Top 5 ranked words by word-topic probability  $\theta$  of each of  $k = 12$  extracted topics of  $n = 831$  documents researching D-ACE excluding documents researching video games. Only the top 5 words were utilized, as word-topic probability  $\theta$  decreased substantially for words ranked lower.

Topics						
Rank	1	2	3	4	5	6
1	Facebook	Socialmedia <sup>a</sup>	Mobile	Video	Read	Digital
2	User	Write	Art	User	Book	Social
3	Social	Draw	Technology	Blog	Student	Technology
4	Motivation	Twitter	Education	Visual	Children	Engage
5	Socialnetwork <sup>a</sup>	Engage	Visual	Engage	School	Storytelling
Topics						
Rank	7	8	9	10	11	12
1	Learn	Music	Interact	Internet	Selfie	Culture
2	Student	Computer	Museum	Online	Image	Media
3	Teach	Technology	Virtual	Social	Photo	Online
4	Education	Song	User	Socialnetwork <sup>a</sup>	Media	Education
5	Motivation	Software	Technology	Behavior	Picture	Behavior

Note: <sup>a</sup> : Social media and social network are fixed terms, that have been grouped together via parsing.