

WHAT CAN WE DO BETTER IN MEDIA ACCESSIBILITY RESEARCH?

AGNIESZKA
SZARKOWSKA

Institute of Applied Linguistics
University of Warsaw

Audiovisual Translation Lab (AVT Lab)

a.szarkowska@uw.edu.pl

@ASzarkowska

<https://avt.ils.uw.edu.pl/>



AVT & MEDIA ACCESSIBILITY RESEARCH

- **Descriptive** research
 - Corpus studies
 - Product analysis
- **Process** research
 - How AVT/MA products are created by translators, subtitlers, audio describers, etc.
- **Reception** research
 - Viewers' reactions and perceptions
 - Solution testing

EXPERIMENTAL METHODS

- **Eye tracking**
- Electroencephalography (EEG)
- Electrodermal activity (EDA)
- Heart rate

PROBLEMS OR WHAT WE CAN DO BETTER

- Research designs
 - Variables
 - Samples
 - Research material
- Data analyses
- Reporting

I. RESEARCH DESIGN

RESEARCH QUESTIONS AND DESIGN

- What is your **research question**?
- How does previous research and literature inform your research question? What are your **hypotheses**?
- How can you **operationalise** your research question? Which eye-tracking measures are appropriate to answer your research question?

PRIMARY AND SECONDARY MEASURES

Primary measures

- Fixation count
- Fixation duration
- Saccade count
- Saccade length
- Pupil dilation

Secondary measures

- Gaze time (dwell time)
- Glance duration
- Time to first fixation
- Revisit (count and duration)
- Skipped subtitles (percentage)
- Word fixation probability (percentage/ratio)

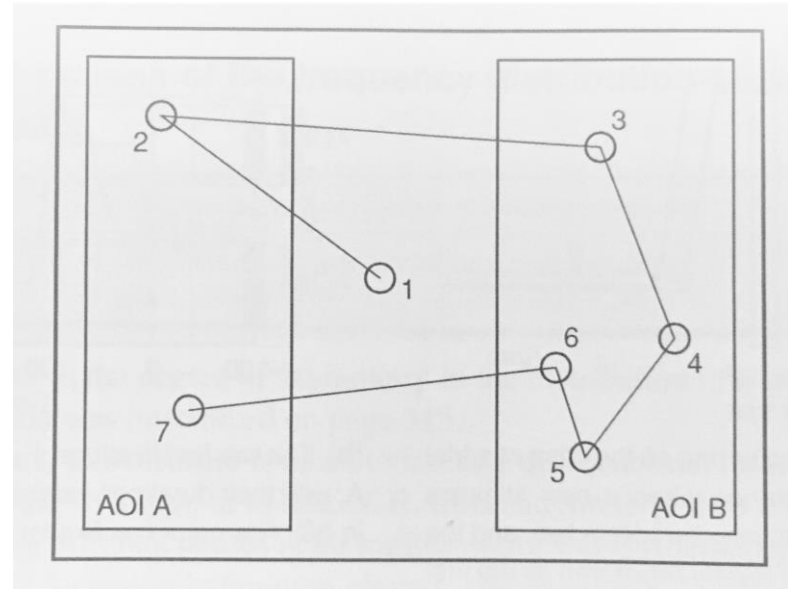
Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In J. Sita, T. Dwyer, S. Redmond, & C. Perkins (Eds.), *Seeing into Screens* (pp. 46-64). Bloomsbury.

EYE-TRACKING METRICS AS DEPENDENT VARIABLES

- If I want to study:
 - time spent on the subtitle → **dwelt time**
 - re-reading of subtitles → **regressions, revisits**
- To facilitate replication:
 - Provide definitions of the metrics
 - Different eye-tracking companies – different names

DWELL TIME

- "One visit in an AOI, from entry to exit"
- Other terms:
 - glance duration
 - gaze duration
 - first-pass fixation time



Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: a comprehensive guide to methods and measures*. Oxford University Press.

WHAT IS YOUR METRIC AND WHY YOU ARE USING IT HERE?

Table 3. Definitions of the eye tracking measures.

Eye tracking measure	Description
Fixation count	The number of fixations in the AOI, averaged per participant per clip. High fixation count can be indicative of higher cognitive effort and/or poorer reading skills.
Mean fixation duration	The duration of a fixation in a subtitle AOI, averaged per clip and per participant. Longer mean fixation duration may be an indication of higher cognitive load.
Absolute reading time	Dwell time, i.e., the sum of durations of all fixations and saccades in an AOI starting with the first fixation. Longer reading time may be related to processing difficulties.

	Polish Voice Over		English Subtitles		Polish Fansubs	
	Mean	SD	Mean	SD	Mean	SD
AOI Dwell Time [ms]	2297.0	1718.6	1797.7	1469.5	426.1	235.4
AOI Gaze Duration	2658.1	1654.1	1624.7	1526.2	753.4	610.3
AOI Glances Count	2.9	1.8	2.4	0.9	2.0	1.0
AOI Revisits	1.9	1.8	1.4	0.9	1.0	1.0
AOI Fixation Count	6.0	4.8	3.5	1.8	1.0	1.0
Fixation Count	381.7	134.6	381.8	183.2	456.2	111.0
AOI Average Fixation Duration [ms]	300.1	92.5	467.0	294.8	161.7	54.4
Average Fixation Duration [ms]	276.7	92.5	271.3	93.6	227.6	55.5
Saccade Count	480.8	63.6	538.2	296.0	607.6	154.8
Saccade Duration Average [ms]	80.9	73.5	49.9	10.5	49.8	3.9
Saccade Amplitude Average [°]	14.2	13.9	8.2	6.5	7.6	2.4
Blink Count	37.9	27.9	72.5	72.1	45.4	26.0

Table 2: Stimulus I: Eye-tracking data overview

DEFINITIONS

Glances count

- Number of glances to a target (saccades coming from outside) within a certain period

Revisits

- Glances count - 1

Dwell time

- Sum of durations from all fixations and saccades that hit the AOI

Gaze duration

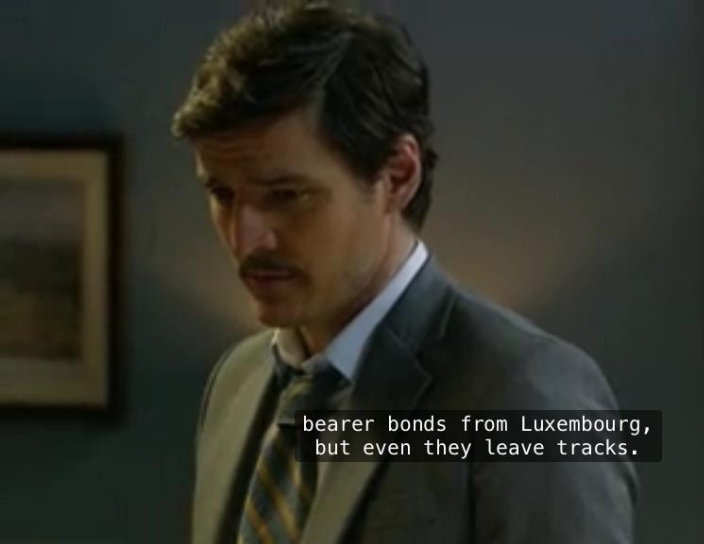
Gaze duration on a word thus contrasts to first fixation duration, the other major reading measure, which is used as an index on word frequency. "Gaze duration" is a reading research term. It is defined exactly as dwell time.

CONSTRUCTING RESEARCH MATERIAL

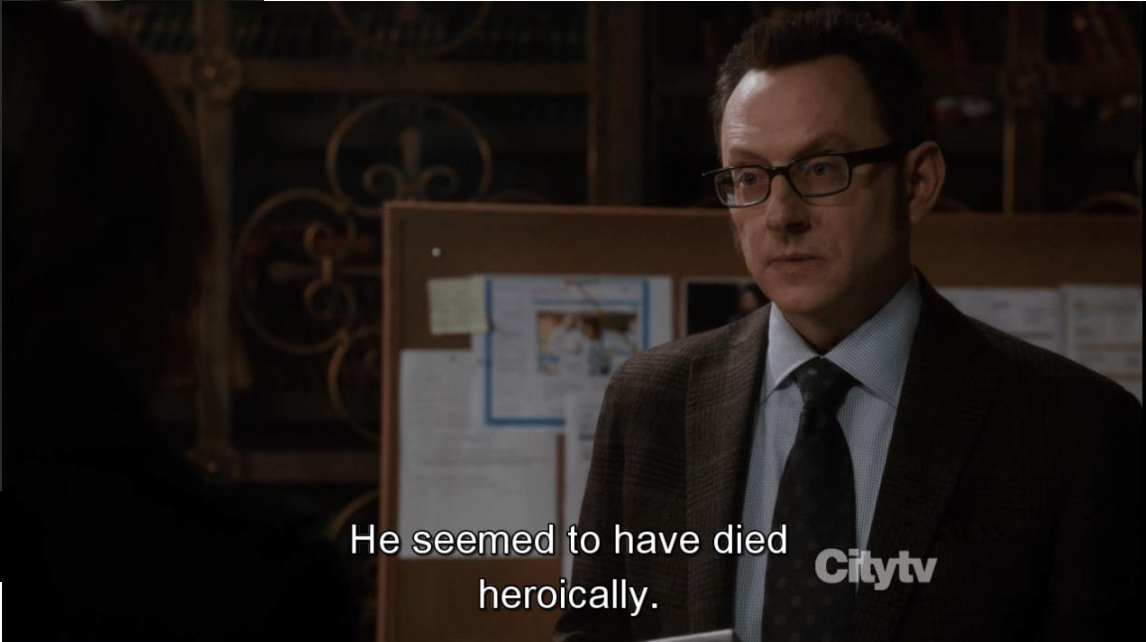


Hopefully, by exposing my very own
medical history, I will inspire
our future leaders to do the same.

SPACE BETWEEN SUBTITLE LINES

A man with dark hair and a mustache, wearing a grey suit and a patterned tie, is looking down and to the left. The background is dark and out of focus.

bearer bonds from Luxembourg,
but even they leave tracks.

A man with glasses, wearing a dark suit and a dark tie, is looking forward. He is standing in front of a bulletin board with various papers pinned to it. The background is dark and out of focus.

He seemed to have died
heroically.

Citytv

AREAS OF INTEREST (AOIs)

You will see the painting
"The Calling of St. Mathew" by Carravaggio.

You will have 15 seconds.

Answer the question that follows.

Press space to continue.



6_problemy 6_z 6_nadwaga
Problemy z nadwagą

6_ma 6_ok 6_100 6 mln 6_Amerykanow
ma ok. 100 mln Amerykanów.

BEYOND AOIs

Data regarding these subtitles were recorded by a software (Eye Scanpath Explorer) specifically designed to simplify the analysis of subtitled material by grouping eye data into two different sets: one for the area **below a threshold line (subtitle region)**, and one for the area above it (**main film zone** or upper area).

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243-272. <https://doi.org/10.1080/15213269.2010.502873>

THRESHOLD LINE

The default position of the threshold line was settled at **675 pixels** from the top of the 1280 x 1024 screen [...].

Fixations occurring **above the threshold line** were considered as fixations on the main film zone, whereas fixations occurring **below** it were considered fixations on the subtitles.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243-272. <https://doi.org/10.1080/15213269.2010.502873>

PIXEL-BASED DEFINITION OF AOI

The fixations were first split according to whether they occurred in the **image area** or the **subtitle area**.

The subtitle area was taken as an area of 1024 (whole width of screen) × 218 pixels that started 50 pixels from the bottom of the display screen.

The subtitle's timing information was used to determine whether a fixation occurred during the presentation of a subtitle.

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(02), 399-418.

<https://doi.org/10.1017/s0142716412000434>

PARTICIPANT PROFILING

- Use established psychometrics to measure:
 - Language proficiency
 - Working memory capacity
 - Viewing habits
 - Immersive tendency
- Asking the right questions in the right way

Scale on light perception, the ability to recognise friends, and to read

1. Which of these best describes your sight with glasses or contact lenses if you normally use them?

A: I have no light perception

B: I can tell by the light where the windows are

C: I can see the shapes of furniture in the room

2. I can recognise a friend by sight alone if . . .

D: I'm close to their face

E: I'm at arm's length away

F: I'm on the other side of the room

G: I'm on the other side of the street

3. I can . . .

H: Read a newspaper headline

I: Read a large print book

J: Read ordinary newsprint

Douglas, G., Corcoran, C., & Pavey, S. (August, 2006). Network 1000: Opinions and circumstances of visually impaired people in Great Britain. Birmingham: University of Birmingham.

Walczak, A., & Fryer, L. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6-17. <https://doi.org/10.1177/0264619616661603>

ASKING ABOUT PARTICIPANTS' AGE

18–30

30–40

40–50

50–60

60 and over

ASKING QUESTIONS

How often do you usually visit the cinema?

Very often _____

Quite often _____

Not very often _____

Not at all _____

How frequently do you usually visit the cinema?

(Please tick whichever category comes closest to the number of times you visit the cinema)

More than once a week _____

Once a week _____

2 or 3 times a month _____

Once a month _____

A few times a year _____

Once a year _____

Less than once a year _____

ASKING QUESTIONS

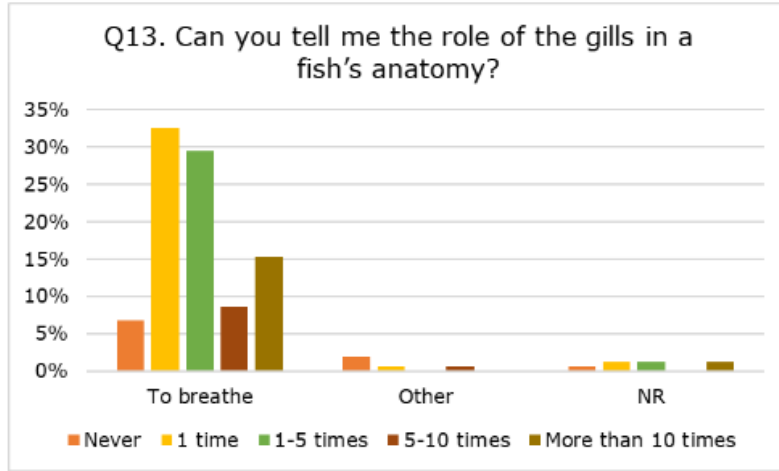


Figure 3. Role of gills in fish's anatomy / Number of viewings

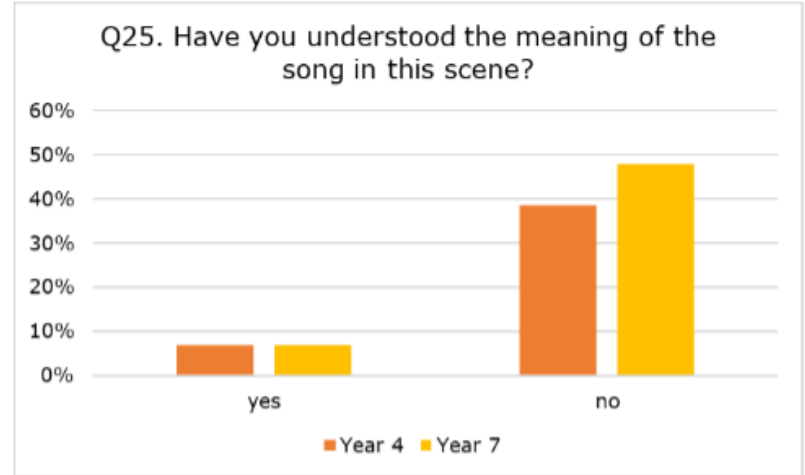


Figure 12. Comprehension of an untranslated song / Year in school

CONFOUNDING VARIABLES IN RESEARCH DESIGN

This study involved a total of sixteen participants (5 male; 11 female) who are native speakers of Japanese. They were split into two groups depending on age category; **Group 1: 18-25 ($n=9$)** and **Group 2: 50+ ($n=7$)**. Experiments were conducted in two locations; whereas the younger age group (control group) was readily available in Ireland where the PI is located, due to the difficulty in finding Japanese subjects in the older age brackets locally in Ireland, this group was recruited in Japan.

II. DATA ANALYSIS

FROM ANOVAs TO MIXED EFFECTS MODELS

- **ANOVA**

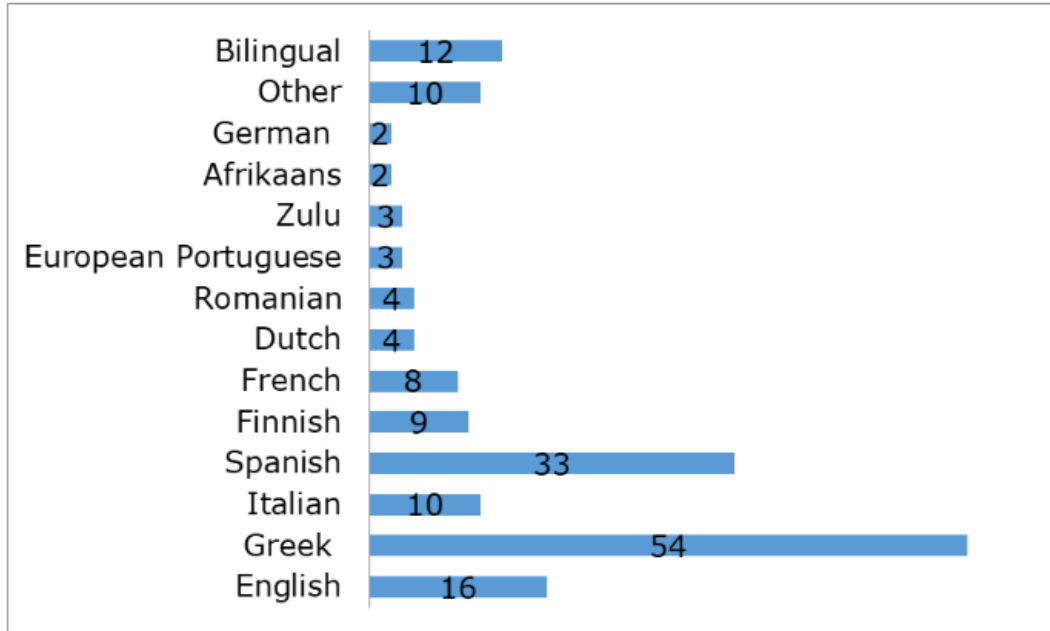
- Mean scores per participant per test
- Traditional ANOVAs disregard the variation of individual scores in test items (particular subtitles)

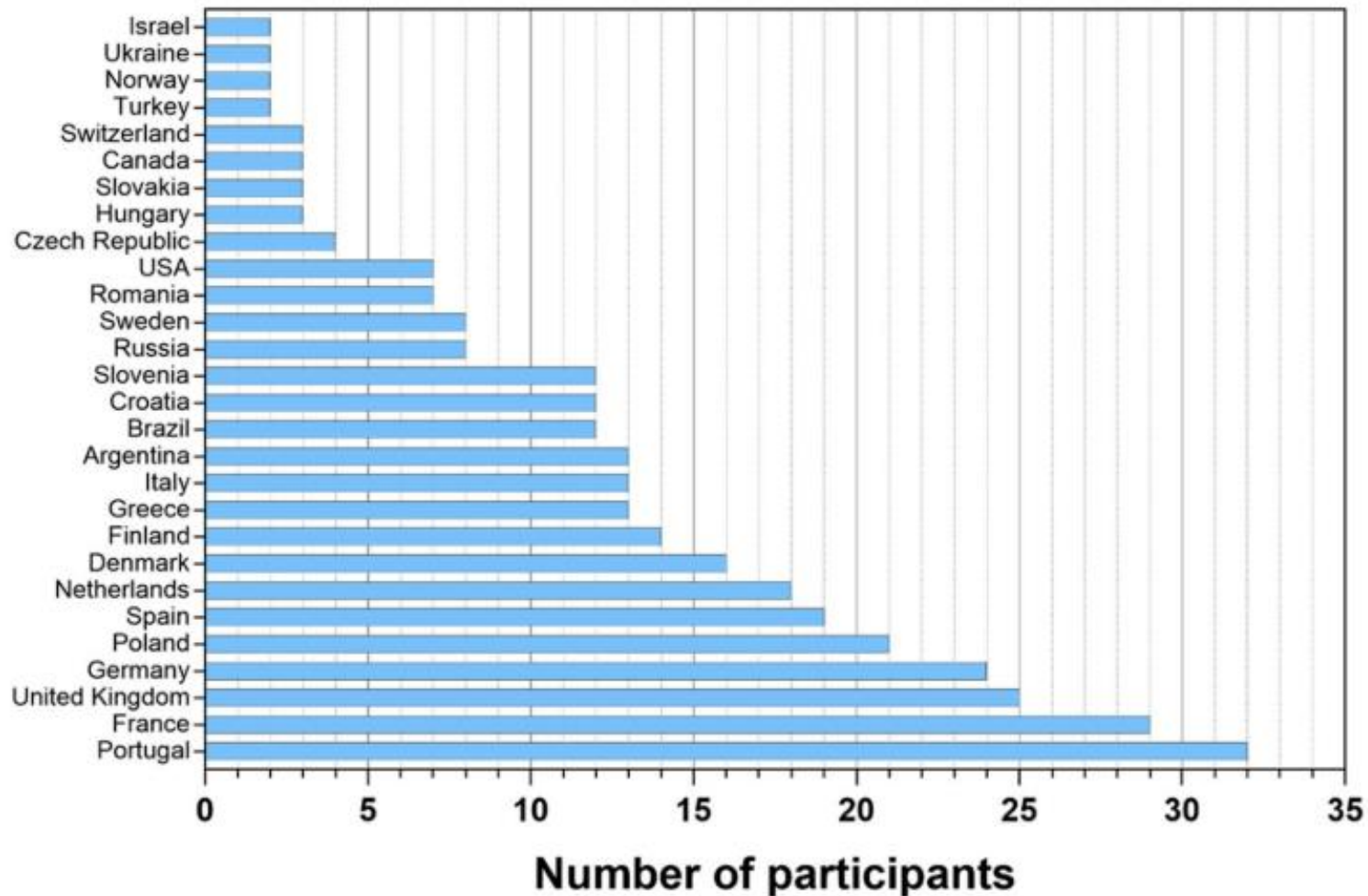
- **Linear Mixed Effects Models**

- Multilevel data which is not fully independent
- Repeated measures designs
- Responses from the same participant are correlated across different linguistic items (subtitles)

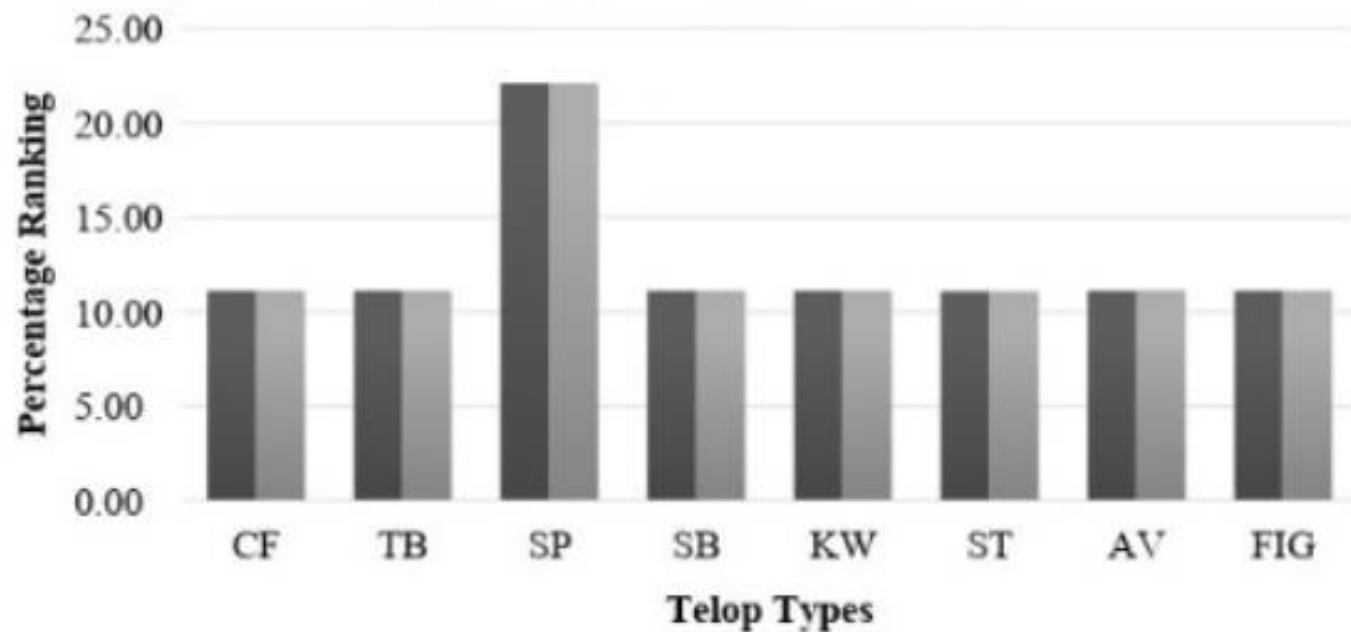
III. REPORTING RESULTS

REPORTING RESULTS



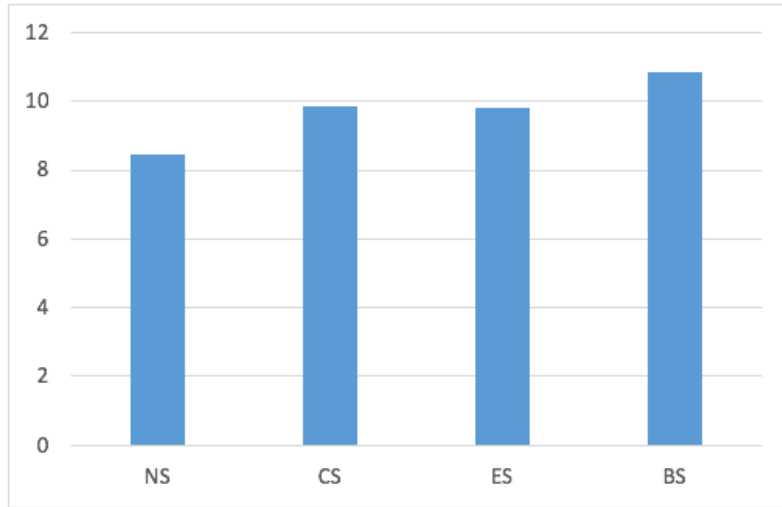


Fixation Duration/Fixation Count



■ Group 1
■ Group 2

REPORTING RESULTS



Video Condition	M (SD)	N
CS	9.88 (5.44)	15
ES	9.82 (4.00)	15
BS	10.83 (5.93)	15
NS	8.45 (3.46)	15

Table 7. Means of recall test scores in different conditions

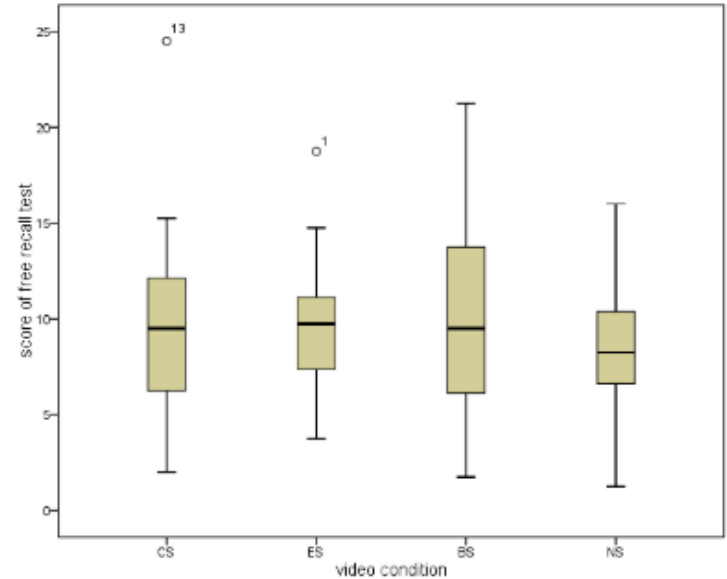
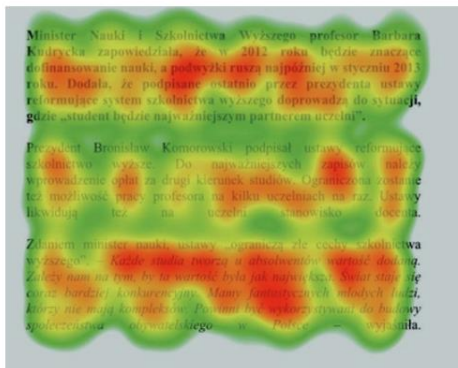


Figure 9. Scores of the free recall test in different video conditions

REPORTING EYE TRACKING RESULTS

ne, określając wcześniej najbardziej istotne dla celu badania parametry/ kryteria: (a) całkowitą liczbę fiksacji, (b) całkowity czas fiksacji, (c) relatywny czas fiksacji:

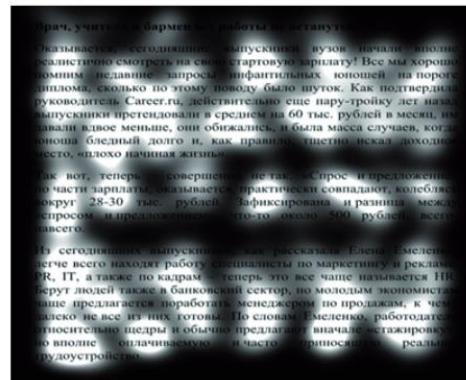


Rys. 2. Kolorowa mapa cieplna uwzględniająca całkowitą liczbę i długość fiksacji podczas tłumaczenia a vista tekstu polskiego na tekst rosyjski.

Mapa pokazuje średnią czasu długości fiksacji dla wszystkich probantów, czyli na jakich elementach najdłużej zatrzymali oni wzrok. Na powyższej mapie widać, iż studenci najdłużej zatrzymywali wzrok na wyrazach „podwyżki”, „ruszą” oraz „studia”. Dowodzi to, że dłużej się zastanawiali nad tymi wyrazami.

wygenerowanie różnych rodzajów map:

(1) Mapy czarno-białe, których stopień transparentności obrazuje natężenie uwagi badanego kierowanej na dany element. Inaczej mówiąc, im jaśniejszy obszar na mapie, tym więcej uwagi badany mu poświęcił, czyli tym dłużej zatrzymał na nim wzrok:



Rys. 1. Czarno-biała mapa cieplna obrazująca długość fiksacji na poszczególnych elementach tekstu podczas tłumaczenia a vista (tłumaczenie tekstu rosyjskiego na tekst polski).

REPORTING

- APA standards
- Descriptive and inferential stats
- Effect sizes

Guideline for Reporting Standards of Eye-tracking Research in Decision Sciences

AUTHORS

Susann Fiedler, Michael Schulte-Mecklenbeck, Frank Renkewitz, Jacob Orquin

AUTHOR ASSERTIONS

Conflict of Interest: No ▾

Public Data: Available ▾

Results

Data quality

Monitoring of data quality during experiment

Proportion of trials excluded for the analysis

Reasons for exclusion

Number of participants excluded from the analysis

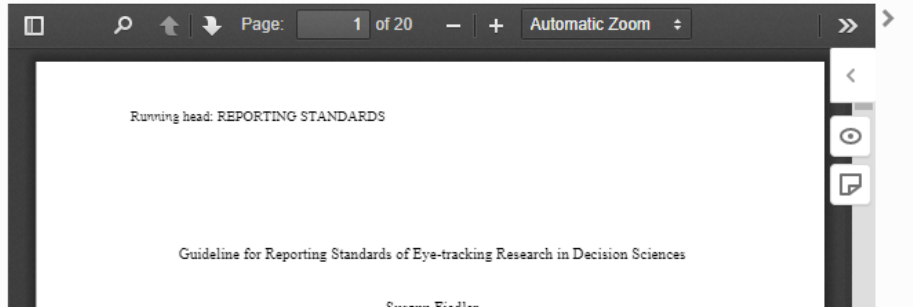
Quality threshold for data exclusion

Percentage of lost data

Dependent measures

Aggregation method for fixations

Additional transformation of the data



<https://psyarxiv.com/f6qcy/>

MIXED METHODS DESIGN

- Triangulation of individual methods to overcome their limitations
- Combining eye tracking with other measures
 - Self-reported cognitive load
 - Comprehension
 - Immersion

BAD RESEARCH?



- Andrew Wakefield *The Lancet*
- *Why Most Published Research Findings Are False?* by John P. A. Ioannidis
- Replication crisis

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associated by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to granuloid ulceration. Histology showed patchy chronic inflammation in 11 of 11 children and reactive ileal lymphoid hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and a possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and normal EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls ($p=0.03$), low haemoglobin in four children, and a low serum IgA in four children.

Interpretation We identify an associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

Lancet 1998; **351**: 637–41

See Commentary page

Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and vomiting and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features, of these children.

Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (including abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for 1 week, accompanied by their parents.

Clinical investigations

We took histories including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria. Developmental assessments included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known causes of childhood neurodegenerative disease. Urinary

DOs

- Have a clear research question
- Choose appropriate measures to help you answer your research question and define them
- Report descriptive AND inferential statistics
- Visualise your data properly
- Data triangulation
- Make your data available in an open access repository

DON'Ts

- Start without a research question (*"I want to do something with eyetracking"* is NOT a good research question)
- Use all available eye-tracking measures available in your software
- Mess up your research design
- Report visualisations or percentages only
- Use eyetracking ONLY
- Report your experiment in a way that's impossible to replicate

AGNIESZKA
SZARKOWSKA



Institute of Applied Linguistics

Audiovisual Translation Lab (AVT Lab)

a.szarkowska@uw.edu.pl

@ASzarkowska

<https://avt.ils.uw.edu.pl/>

<https://www.facebook.com/AVTlab/>