Baron-Cohen et al. (2001): The Eyes Test

Lesson 1.2.2 | Concepts and methods

1. Psychology and aims

Theory of mind: attributing mental states to oneself and others. **Mentalising** includes inferring beliefs, intentions and feelings. **Social sensitivity** involves noticing subtle mental-state cues. These are cognitive processes; one task does not measure every aspect of empathy.

The 1997 test had ceiling and discrimination problems. A revised test aimed to detect subtler differences, compare autistic adults with comparison groups, examine sex differences and test the relationship with Autism-Spectrum Quotient (AQ) scores. The paper used the historical diagnostic terms AS/HFA.

2. Improvements to the test

Twenty-five photographs became 36 scored items; two choices became four (chance success fell from 50% to 25%). Complex mental states, closer distractors, balanced male/female faces and removal of gaze-direction shortcuts improved discrimination. A glossary reduced vocabulary confusion.

3. Participants

Group	Sample
1: historical AS/HFA group	15 adult men, support-group adverts; mean age 29.7, mean IQ 115.
2: community comparison	122 adults, classes/library users; mean age 46.5 for 88 with age data.
3: students	103 Cambridge undergraduates: 53 men, 50 women; mean age 20.8.
4: IQ-matched comparison	14 adults; mean age 28.0, mean IQ 116.

4. Design and procedure

Independent-group comparisons used naturally occurring diagnosis and sex, not randomly assigned traits. The original paper describes individual quiet-room testing with a glossary. Participants selected the best mental-state word for each photograph of the eye region. Groups 1, 3 and 4 completed AQ; Group 1 also judged the faces' sex.

Baron-Cohen et al. (2001): The Eyes Test

Lesson 1.2.2 | Evidence and interpretation

1. Test construction and measures

Eight pilot judges checked target words and distractors. Larger item analysis removed four of 40 candidate photographs, leaving 36. Each item offered four words; one target counted as correct. AQ is a self-report trait measure scored 0-50, distinct from the performance test.

2. Eyes Test scores

Group	Mean / 36	Standard deviation
1: AS/HFA	21.9	6.6
2: community	26.2	3.6
3: students	28.0	3.5
4: IQ-matched	30.9	3.0

The historical AS/HFA group scored lower on average than the comparison groups. Members of every group varied; mean differences do not classify every individual.

3. AQ and other findings

AQ group	Usable n	Mean / 50
1: AS/HFA	14	34.4
3: students	79	18.3
4: IQ-matched	14	18.9

Higher AQ scores tended to accompany lower Eyes scores; the combined correlation was **-0.53**. IQ was not significantly correlated with Eyes performance in the available measured sample.

Female Eyes means were higher, but this was a non-significant trend. Student AQ means were 19.5 for men and 16.6 for women, a reported significant difference. Every AS/HFA participant scored at least 33/36 on sex recognition.

Conclusion: the revised task detected subtle social-cognitive differences. Successful sex recognition alongside harder mental-state judgements argues against a simple inability to process eye photographs, but does not measure all everyday social understanding.

Baron-Cohen et al. (2001): The Eyes Test

Lesson 1.2.2 | Evaluation and connections

1. Strengths

Systematic revision: pilot criteria and item analysis improved sensitivity and reduced easy shortcuts.
Comparisons: the IQ-matched group addressed general intelligence as an alternative explanation.
Glossary and control task: reduced vocabulary misunderstanding and checked basic processing of the same stimuli.

2. Limitations

Ecological validity: static eye photographs and four supplied words omit voice, movement, body language and social context.
Generalisation: the AS/HFA group contained only 15 men, limiting transfer across sex, age and ability.
Construct validity: language, cultural knowledge and consensus-selected target words may influence scores. This is one component of mentalising, not a complete empathy assessment.

3. Ethics and interpretation

Diagnosis and social-cognition scores are sensitive; voluntary consent, respectful treatment and protection of identifiable data matter. Do not invent modern safeguards not described in the report. A group difference should not be framed as a judgement of a person's worth or entire social ability.

4. Application and debates

Application: the task can support research on social sensitivity; it is not a stand-alone diagnosis and no treatment was tested.
Individual/situational: cognitive characteristics differ, but restricted cues and word choices also shape performance.
Nature/nurture: diagnostic comparisons did not isolate genes, maturation or experience as the cause.

> Three distinctions: Eyes score is not AQ; a negative correlation is not causation; a female mean advantage was not a significant Eyes Test sex difference here.

Hint: Keep test sample sizes separate from usable AQ sample sizes. For procedure questions, the original study describes individual quiet-room testing rather than an online administration.