

Eksamensoppgave i PSY2022 Forskningsdesign

Faglig kontakt under eksamen: Ute Gabriel

Tlf.: 73 59 19 60

Eksamensdato: 02.06.2015

Eksamenstid (fra-til): 09.00-13:00

Hjelpemiddelkode/Tillatte hjelpemidler: Ingen

Annen informasjon:

Målform/språk: Bokmål

Antall sider: 5

Antall sider vedlegg: 1

Kontrollert av:

Eksamen består av elleve oppgaver. Alle oppgaver skal besvares. Oppgavene vektes i samsvar med %-poeng som er oppgitt bak hver oppgave (Oppgave 1 til 9 utgjør til sammen 30%, Oppgave 10 utgjør 35% og Oppgave 11 utgjør 35%, Totalt: 100%).

På oppgavene 1, 2 og 7 må du enten svare på oppgavearket og levere inn dette – eller du kan skrive oppgavens nummer og riktig svaralternativ på svararket.

Oppgave 1. En forsker formulerer en hypotese om at det vil være en betydelig forskjell bekymringene uttrykt av gravide i løpet av første, andre eller tredje trimes svangerskapet. Hva er den uavhengige variabelen i denne hypotesen? (0.5

- (a) Forskjell i bekymringene
- (b) Uttrykk av bekymringer
- (c) Gravide
- (d) Trimester av svangerskapet

Oppgave 2. To observatører har separat skåret et barns lek-atferd for aggresjon. Hvilket målebegrep indikerer graden av enighet mellom de to observatørene? (0.5%)

- (a) Inter-rater reliabilitet
- (b) Face validitet
- (c) Test-retest reliabilitet
- (d) Kriterievaliditet

Oppgave 3. Hva er “personopplysninger”? (0.5%)

Oppgave 4. Gi et eksempel på “indirekte personopplysninger”: (1%)

Oppgave 5. Hva er forskjellen mellom “konfidensialitet” og “anonymitet” i forskningssammenheng? (2%)

Oppgave 6. Hva menes med at en person gir sitt “*frivillige, uttrykkelige og informerte samtykke*”? (4.5%)

Oppgave 7. Angi for hvert av de følgende utsagn om det er korrekt eller feil. (2%)

Forutsatt at alt annet er likt, gjelder ...	Korrekt	Feil
7.1 Når effektstørrelsen øker, synker statistisk styrke (power)	<input type="checkbox"/>	<input type="checkbox"/>
7.2 Når utvalgsstørrelsen øker, øker statistisk styrke (power).	<input type="checkbox"/>	<input type="checkbox"/>
7.3 Når Type I feil (alpha) øker, synker statistisk styrke (power).	<input type="checkbox"/>	<input type="checkbox"/>
7.4 Type II feil (beta) er ikke relatert til statistisk styrke (power).	<input type="checkbox"/>	<input type="checkbox"/>

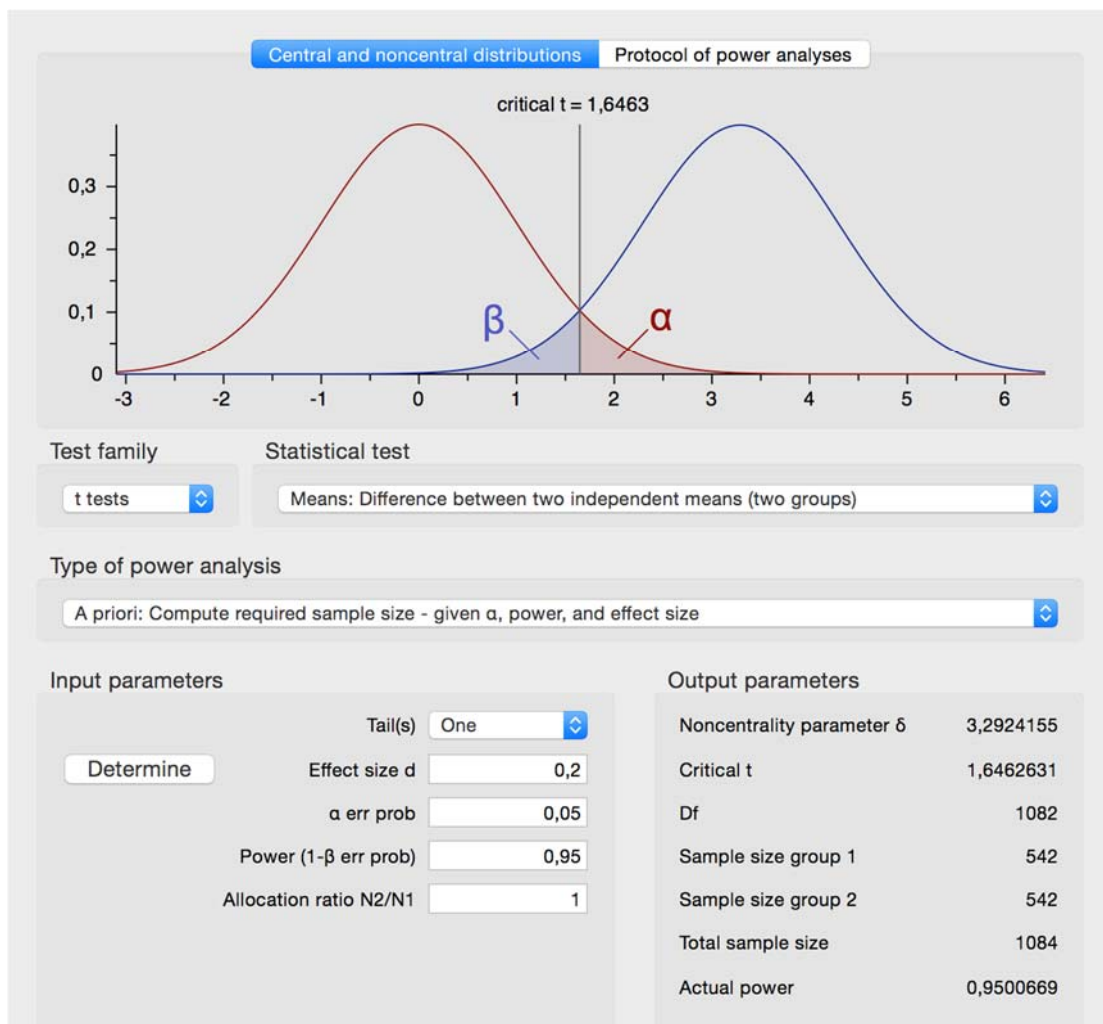
Oppgave 8. I forberedelse til en studie gjennomførte en forsker en *a priori* statistisk styrkeanalyse (*a priori power analysis*). Resultatet fra den analysen vises i Figur 1. (13%)

8.1 Hva var det som forskeren *ønsket* å finne ut ved å gjennomføre en *a priori* statistisk styrkeanalyse (*a priori power analysis*)?

8.2 Hva er svaret på det som forskeren *ønsket* å finne ut ved å gjennomføre en *a priori* statistisk styrke analyse (*a priori power analysis*)?

8.3 Hvilke parametre måtte forskeren angi og velge for å kunne gjennomføre analysen?

8.4 Hva kan vi lære om forskerens forskningsdesign fra ”input parameters” i Figur 1?



Figur 1: Resultatet fra en *a priori* statistisk styrkeanalyse (Oppgave 8)

Oppgave 9. Nedenfor vises et utdrag fra et spørreskjema som ble utviklet for å undersøke narkotikamisbruk blant norske elever i videregående skole. Identifiser mulige svakheter og problemer i utvalg og formulering av spørsmål og svar, og foreslå endringer. (Forslag som går på Lay-out er ikke av interesse her.) (6%)

a) Hva er navnet ditt? _____

b) Hva er navnet til skolen din? _____

c) Hva er ditt kjønn? Kvinne Mann

d) Hva er fødselsdagen din? Dag_____ Måned _____ År_____

e) Hvor ofte konsumerer du tobakk, alkohol, marihuana, ecstasy, amphetamine, heroin og kokain?

Aldri Sjelden Ofte Alltid

f) Hvor ofte i løpet av den siste måned har du hatt ubeskyttet sex?

0 1-2 2-4 3-6 Flere

g) De fleste unge prøver narkotika ut av gruppepress. Hvor enig eller uenig er du i at dine beste venner er ansvarlig for at du prøver narkotika?

ganske uenig
 delvis uenig
 vet ikke
 delvis enig
 ganske enig

Oppgave 10. Ledelsen ved en norsk videregående skole ønsker å vite hvorfor ett økende antall av elevene deres prøver narkotiske stoffer. Som rådgiver ved skolen, blir du bedt om å designe en studie for å undersøke dette. (35%, lik fordelt mellom alle fire deloppgaver)

10.1: Bestem deg for en forskningsmetode, og forklar hvorfor du velger den metoden.

10.2: Hvilke etiske bekymringer bør du være klar over, og hvordan ville du håndtere dem?

10.3: Formuler minst tre hypoteser for utredningen.

10.4: Forklar ditt utvalgsstrategi (sampling strategy).

Oppgave 11. Følgende spørsmål gjelder Eksperiment 3 i Bem (2011), beskrevet i vedlegg (Appendiks).

11.1: Spesifiser det eksperimentelle designet: (20%)

- a. Hva er de/n eksperimentelt manipulererte variabelen/e?
- b. Hva er de/n avhengige variabelen/e?
- c. Oppbygging av eksperimentet:
Hva er “*blocks*” i dette eksperimentet?
Hvor mange *blocks* er det?

Hva er “*trials*” i dette eksperimentet?
Hvor mange *trials* er det?

Hva er “*events*” i dette eksperimentet?
Hvor mange *events* er det?
- d. Har eksperimentet et between-subjects eller en within-subject design? Gi en kort begrunnelse for svaret ditt.
- e. Hvordan ble randomisering implementert i dette eksperimentet?

11.2: Intern og ekstern validitet (15%)

- a. Hva betegnes som intern validitet?
- b. Hva betegnes som ekstern validitet?
- c. Vurder den interne og eksterne validiteten i dette eksperiment.

FEELING THE FUTURE

Retroactive Priming

Experiment 3: Retroactive Priming I

In recent years, priming experiments have become a staple of cognitive and cognitive social psychology (Bargh & Ferguson, 2000; Fazio, 2001; Klauer & Musch, 2003). In a typical affective priming experiment, participants are asked to judge as quickly as they can whether a picture is pleasant or unpleasant, and their response time is measured. Just before the picture appears, a positive or negative word (e.g., *beautiful*, *ugly*) is flashed briefly on the screen; this word is called the prime. Individuals typically respond more quickly when the valences of the prime and the picture are congruent (both are positive or both are negative) than when they are incongruent. In our retroactive version of the procedure, the prime appeared after rather than before participants made their judgments of the pictures.

Because slower responding on congruent trials than on incongruent trials— called a contrast effect— has also been observed in some priming experiments (Hermans, Spruyt, De Houwer, & Eelen, 2003; Klauer, Teige-Mocigemba, & Spruyt, 2009), we also ran a standard nonretroactive priming procedure in each session to ensure that our protocol would produce the usual (noncontrast) priming effect (see also de Boer & Bierman, 2006). Because this turned out to be the case, the psi hypothesis was that the retroactive procedure would also produce faster responding on congruent trials than on incongruent trials.

Table 2
Four Analyses of Precognitive Avoidance of Negative Stimuli

Hit rate % across participants	Binomial test across trials	Phi coefficient across participants	Hit rate % using an empirical baseline
51,7%	2,790/5,400 = 51,7%	.034	51,7%
$t(149) = 2.39$	$z = 2.44$	$t(149) = 2.46$	$t(149) = 2.37$
$p = .009$	$p = 2.44$	$p = 0.008$	$p = 0.10$
$d = 0.20$	$r = .20$	$d = 0.20$	$d = 0.19$

Method

One hundred Cornell undergraduates, 69 women and 31 men, participated in a 15- to 20-min experiment. They were shown a picture on each of 64 trials and were asked to press one of two keys on the keyboard as quickly as they could to indicate whether the picture was pleasant or unpleasant. The participant's response time in making this judgment was the dependent variable, and the difference in mean response times between incongruent and congruent trials is the index of a priming effect, with positive differences denoting faster responding on congruent trials.

The first 32 trials constituted the retroactive priming procedure, and participants were told that a word would be flashed on the screen just after they made their judgment of the picture. The remaining 32 trials constituted the standard forward priming procedure, and participants were told that "from this point on, the flashed word will appear before rather than after you have made your response." Prior to beginning the actual experimental procedure, participants responded to the two items on the stimulus seeking scale and then had the same 3-min relaxation period described in the previous experiments.

The pictures were again drawn from the IAPS set and were randomly assigned to the forward and retroactive sections of the protocol, with the restriction that an equal number of positive and

negative pictures appear in each section. The same 16 positive and 16 negative prime words appeared in both sections, and a prime was randomly selected on each trial before the picture was presented (in the forward priming procedure) or after the participant had responded to the picture (in the retroactive priming procedure). As a result, congruent trials and incongruent trials were randomly sequenced and did not necessarily occur in equal numbers. This made it virtually impossible for participants to anticipate the type of trial coming up by knowing the types of trials that had already occurred. In this experiment, randomizing was implemented by Marsaglia's PRNG algorithm.

Figure 2 displays the time sequence of events for the forward priming and retroactive priming trials, respectively. In both procedures, there was a 2,000-ms interval between trials during which a Hubble photograph of the starry sky appeared on the screen.

Results and Discussion

Several methods for analyzing response-time data from priming experiments have evolved over the years (Ratcliff, 1993). First, trials on which a participant makes an error in judging the picture to be pleasant or unpleasant are excluded from the analysis. In the present experiment, the median number of errors was three out of 64 trials, and the data from three participants were discarded because they made errors on 16 (25%) or more of the trials, reducing the number of participants to 97. Second, because response-time data are positively skewed, each response time (RT) is usually transformed prior to analysis using either an inverse transformation ($1/RT$) or a log transformation ($\log RT$). Finally, trials yielding very short or very long response times are considered to be spurious outliers and are excluded from the analysis. Ratcliff (1993) suggested using more than one cutoff criterion to ensure "that an effect is significant over some range of nonextreme cutoffs" (p. 519). Accordingly, Table 4 presents four analyses, using both data transformations and two different cutoff criteria for long response times, 1,500 ms and 2,500 ms. The first criterion excludes 3.1% of the trials; the second excludes 0.5% of the trials. As shown in the table, the standard forward priming procedure produced the usual result. For example, with a 1,500-ms cutoff criterion and the inverse transformation, participants were 23.6 ms faster on congruent trials than on incongruent trials, $t(96) = 4.91$, $p < .00001$, $d = 0.45$. The retroactive procedure also yielded the predicted psi effect: With a 1,500-ms cutoff criterion and the inverse transformation, participants were 15.0 ms faster on congruent trials than on incongruent trials, $t(96) = 2.55$, $p = .006$, $d = 0.25$. The results were consistent across the range defined by the two cutoff criteria and under both data transformations. To provide an analysis that avoids distribution assumptions, Table 4 also displays the percentage of participants who had positive priming scores in each condition, evaluated by an exact binomial test. For example, with a 1,500-ms outlier cutoff criterion, 64.9% of participants produced a positive forward priming effect ($p = .002$), and 60.8% produced a positive retroactive priming effect ($p = .021$). There was no significant correlation between stimulus seeking and either priming effect ($r = .02$ and $r = .05$ for the forward and retroactive priming effects, respectively).

Eksamen PSY2022 (Våren)

SENSORVEILEDNING

Eksamen består av elleve oppgaver. Alle oppgaver skal besvares. Oppgavene vektes i samsvar med %-poeng som er oppgitt bak hver oppgave (Oppgave 1 til 9 utgjør til sammen 30%, Oppgave 10 utgjør 35% og Oppgave 11 utgjør 35%, Totalt: 100%).

Oppgave 1. En forsker formulerer en hypotese om at det vil være en betydelig forskjell i bekymringene uttrykt av gravide i løpet av første, andre eller tredje trimester av svangerskapet. Hva er den uavhengige variabelen i denne hypotesen? (0.5%)

- (a) Forskjell i bekymringene
- (b) Uttrykk av bekymringer
- (c) Gravide
- (d) Trimester av svangerskapet

Answer is 'd'.

The researcher is examining whether the trimester of pregnancy affects expressed concerns. Trimester of pregnancy is thus the independent variable, and expressed concerns is dependent variable; Difference in concerns refers to changes in the dependent variable; pregnant women are the participants.

Oppgave 2. To observatører har separat skåret et barns lek-atferd for aggresjon. Hvilket målebegrep indikerer graden av enighet mellom de to observatørene? (0.5%)

- (a) Inter-rater reliabilitet
- (b) Face validitet
- (c) Test-retest reliabilitet
- (d) Kriterievaliditet

Answer is 'a'.

The level of agreement between the two observers is an issue of reliability, not validity. Test-retest reliability refers to the consistency between an individual's scores on the same test taken at two or more different times, whereas inter-rater reliability refers to the degree of agreement between different observers or raters.

Oppgave 3. Hva er "personopplysninger"? (0.5%)

=> opplysninger som direkte eller indirekte kan knyttes til en enkeltperson.

Oppgave 4. Gi et eksempel på "indirekte personopplysninger": (1%)

- *Det finnes i prinsippet to mulige eksempler (studenten bør presentere ET av de):*
- *opplysninger som registreres med et referansenummer og viser til en atskilt liste med for ek-*

sempel navn eller personnummer

- *bakgrunnsopplysninger som fører til at personen kan identifiseres (f.eks. kombinasjon av kjønn, stilling, alder som identifiserer "den ene" mannlige 50-årige postdoc som finnes i Trondheim)*

Oppgave 5. Hva er forskjellen mellom "konfidensialitet" og "anonymitet" i forskningssammenheng? (2%)

"konfidensialitet" handler om personidentifiserbare opplysninger => forsker må hindre bruk og formidling av informasjon som kan skade enkeltpersonene det forsakes på.

vs.

*"anonymitet" handler om opplysningene som kan **ikke** knyttes til en enkeltperson (anonymiserte data er data der alle personentydige kjennetegn er fjernet)*

Oppgave 6. Hva menes med at en person gir sitt "frivillige, uttrykkelige og informerte samtykke"? (4.5%)

- ⇒ *frivillig: person decides voluntarily – no press (neither direct nor indirect) to participate and may end participation at any time*
- ⇒ *uttrykkelig: active consent (in contrast to "opt out")*
- ⇒ *informert: in general: purpose of the study, what participation involves (probable risks involved), what happens to the data; information that participation is voluntary; contact details; (additional information – depends on the kind of study ...)*

Oppgave 7. Angi for hvert av de følgende utsagn om det er korrekt eller feil. (2%)

	Forutsatt at alt annet er likt, gjelder ...	Korrekt	Feil
7.1	Når effektstørrelsen øker, synker statistisk styrke (power)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7.2	Når utvalgsstørrelsen øker, øker statistisk styrke (power).	<input checked="" type="checkbox"/>	<input type="checkbox"/>
7.3	Når Type I feil (alpha) øker, synker statistisk styrke (power).	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7.4	Type II feil (beta) er ikke relatert til statistisk styrke (power).	<input type="checkbox"/>	<input checked="" type="checkbox"/>

=> Each correct answer counts 0.5%

Oppgave 8. I forberedelse til en studie gjennomførte en forsker en *a priori* statistisk styrkeanalyse (*a priori power analysis*). Resultatet fra den analysen vises i Figur 1. (13%)

8.1 Hva var det som forskeren *ønsket* å finne ut ved å gjennomføre en *a priori* statistisk styrkeanalyse (*a priori power analysis*)?

=> *Appropriate sample size (given in Figure 1)*

8.2 Hva er svaret på det som forskeren *ønsket* å finne ut ved å gjennomføre en *a priori* statistisk styrke analyse (*a priori power analysis*)?

=> *Figure 1: total sample size of 1084 participants (or 542 per group – two groups)*

8.3 Hvilke parametre måtte forskeren angi og velge for å kunne gjennomføre analysen?

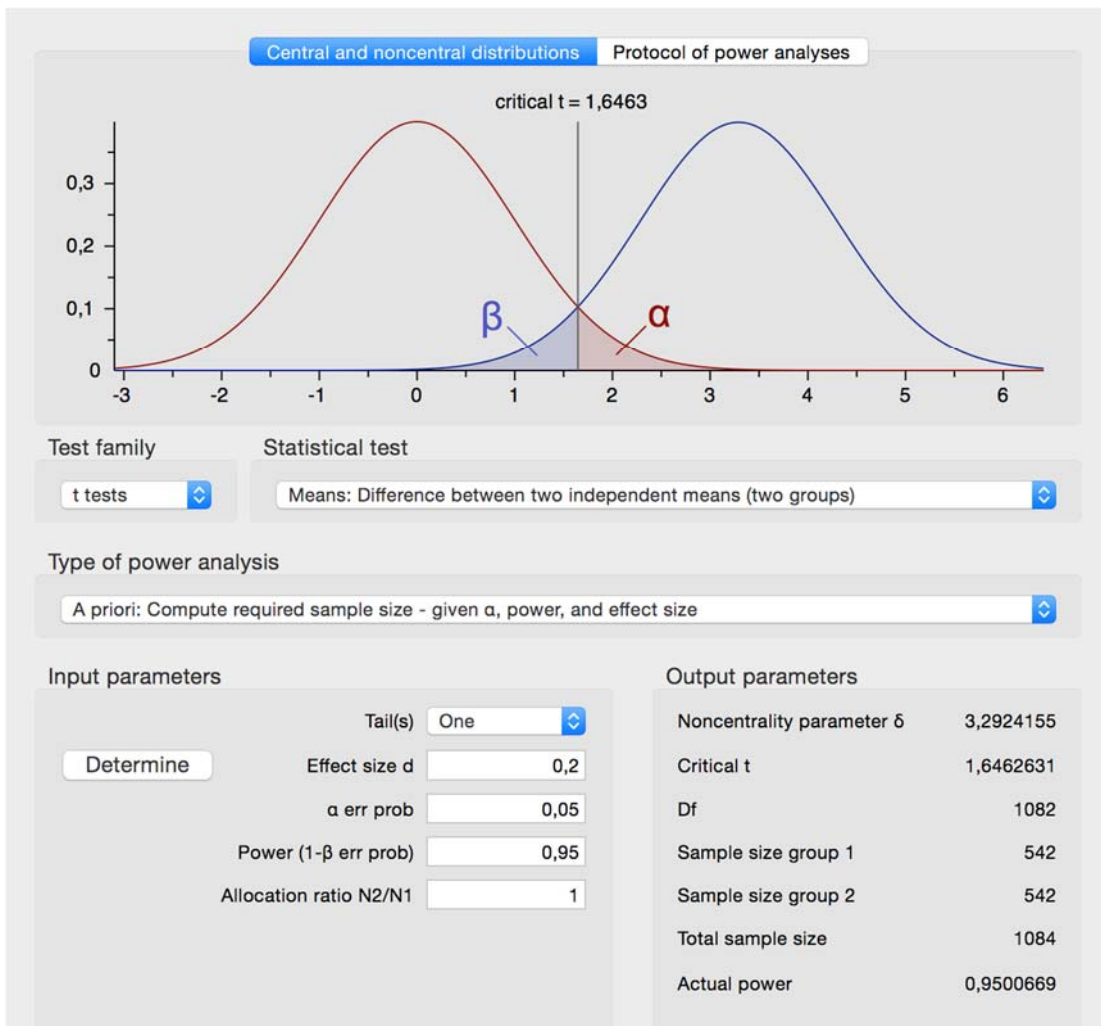
=> (*a priori power analyse*) *t-test, independent samples; $n1 = n2$; one tailed*

=> *alpha, power, effect size*

(*this is all given in Figure 1*)

8.4 Hva kan vi lære om forskerens forskningsdesign fra ”input parameters” i Figur 1?

- ⇒ *is going to compare the mean from two independent samples (t-test, independent samples)*
- ⇒ *plans equal sample sizes ($N1/N2$ ratio)*
- ⇒ *has a directed hypothesis (one tailed)*
- ⇒ *standard choice of alpha (5%)*
- ⇒ *but: beta = 5% => thinks that the consequences of a beta error (falsely rejecting $H0$) is as severe as the consequences of an alpha error (falsely accepting $H0$); results in power 95% (probability to correctly reject $H0$); standard choice of beta = 4x alpha.*



Figur 1: Resultatet fra en a priori statistisk styrkeanalyse (Oppgave 8)

Oppgave 9. Nedenfor vises et utdrag fra et spørreskjema som ble utviklet for å undersøke narkotikamisbruk blant norske elever i videregående skole. Identifiser mulige svakheter og problemer i utvalg og formulering av spørsmål og svar, og foreslå endringer. (Forslag som går på Lay-out er ikke av interesse her.) (6%)

- Hva er navnet ditt? _____
- Hva er navnet til skolen din? _____
- Hva er ditt kjønn? Kvinne Mann
- Hva er fødselsdagen din? Dag _____ Måned _____ År _____

e) Hvor ofte konsumerer du tobakk, alkohol, marihuana, ecstasy, amphetamine, heroin og kokain?

- Aldri Sjelden Ofte Alltid

f) Hvor ofte i løpet av den siste måned har du hatt ubeskyttet sex?

- 0 1-2 2-4 3-6 Flere

g) De fleste unge prøver narkotika ut av gruppepress. Hvor enig eller uenig er du i at dine beste venner er ansvarlig for at du prøver narkotika?

- ganske uenig
 delvis uenig
 vet ikke
 delvis enig
 ganske enig

The questions from a) to d) are considered sensitive. It is possible to track down respondents by this personal information. To ensure anonymity of the respondents, a) needs to be dropped, and d) needs to be reformulated by simply asking the respondent's age.

Question e) is asking the frequency of tobacco, alcohol and drug use (three behavior types) in one question. A separate question needs to be formulated for each behavior type. For example,

	<i>Aldri</i>	<i>Sjelden</i>	<i>Ofte</i>	<i>Alltid</i>
<i>Hvor ofte konsumerer du tobakk?</i>	<i>O</i>	<i>O</i>	<i>O</i>	<i>O</i>
<i>Hvor ofte konsumerer du alkohol?</i>	<i>O</i>	<i>O</i>	<i>O</i>	<i>O</i>
<i>Hvor ofte konsumerer du marihuana, ecstasy, amfetamin, heroin eller kokain?</i>	<i>O</i>	<i>O</i>	<i>O</i>	<i>O</i>

Note: The last one can further be separated into questions asking the use of each specific drug.

The overlapping behavioral frequency specified in answering options (i.e. 0, 1-2, 2-4, 3-6, flere) for question f) should be made into mutually exclusive (e.g. 0, 1-2, 3-4, 5-6, flere) options.

Question g) begins with specifically charged words, i.e., it emphasizes the role of group pressure. This pushes the respondents in a particular direction when they answer if their best friends are responsible for their behavior. Therefore, the question g) is better formulated by dropping the leading sentence in the beginning.

The answering option specified for question g) – ‘vet ikke’ – should either be dropped or placed in the beginning/end of the answering options.

Oppgave 10. Ledelsen ved en norsk videregående skole ønsker å vite hvorfor ett økende antall av elevene deres prøver narkotiske stoffer. Som rådgiver ved skolen, blir du bedt om å designe en studie for å undersøke dette. (35%, lik fordelt mellom alle fire deloppgaver)

10.5: Bestem deg for en forskningsmetode, og forklar hvorfor du velger den metoden.

Candidate should demonstrate their understanding of experimental designs, qualitative methods (participant observations, case studies, open ended interview and focus groups), content analysis and archival research, and quantitative methods (survey) by identifying an appropriate method for studying increased drug use among students. Candidate's explanation of the chosen method would reveal the level of critical thinking and judgment.

10.6: Hvilke etiske bekymringer bør du være klar over, og hvordan ville du håndtere dem?

Candidate should be aware of the fact that the study population is not adults. In addition, the behavior under investigation is considered socially undesirable. Hence the core codes of ethics, i.e. participants should not intentionally be physically or mentally harmed and their right to privacy must be respected, should be followed with utmost care. As such, confidentiality, anonymity, informed consent, and consent from guardians must be carefully addressed.

10.7: Formuler minst tre hypoteser for utredningen.

Candidates should demonstrate that they are capable of constructing research questions and formulating testable hypotheses. The hypotheses should spell out the potential association between measurable variables, e.g., frequency of drug abuse, peer pressure.

10.8: Forklar ditt utvalgsstrategi (sampling strategy).

Depending on the chosen research method, the candidate should explain the reason for choosing a specific sampling method. The candidate's judgment of the chosen methods should reflect the advantages and disadvantages of probability and nonprobability sampling methods.

Oppgave 11. Følgende spørsmål gjelder Eksperiment 3 i Bem (2011), beskrevet i vedlegg (Appendiks).

11.1: Spesifiser det eksperimentelle designet: (20%)

a. Hva er de/n eksperimentelt manipulerte variabelen/e?

*Manip. V1: Retroactive priming/standard forward priming
Manip. V2: Congruent/incongruent trials*

b. Hva er de/n avhengige variabelen/e?

IV: response times (per trial)

c. Oppbygging av eksperimentet:

Hva er “*blocks*” i dette eksperimentet?
Hvor mange *blocks* er det?

Retroactive/standard forward = 2 blocks

Hva er “*trials*” i dette eksperimentet?
Hvor mange *trials* er det?

In the Method section, first paragraph states that participants are presented 64 trials. The second paragraph states that 32 trials were presented in Retroactive priming and 32 in Standard Priming.

(Among the 32 trials in each block, the prime was randomly selected to each picture, so that the number of congruent and incongruent trials varies.)

The trial line is presented in the Appendix in Fig. 2.

Hva er “*events*” i dette eksperimentet?
Hvor mange *events* er det?

The events are labeled in the Appendix in Fig. 2 – 6 events per trial or 5 events per trial if the 2000ms presentation of a Hubble photograph is considered as “inter-trial-interval”.

d. Har eksperimentet et between-subjects eller en within-subject design? Gi en kort begrunnelse for svaret ditt.

Within-subject design; all participants receive all levels of the two manipulated variables.

e. Hvordan ble randomisering implementert i dette eksperimentet?

Random assignment of pictures to blocks (retroactive/standard forward); random sequence of congruent and incongruent trials per block;

11.2: Intern og ekstern validitet (15%)

a. Hva betegnes som intern validitet?

“Internal validity refers to the confidence with which we can draw cause-and-effect conclusions from our research results. To what extent are we certain that the independent variable, or treatment, manipulated by the experimenter is the sole source or cause of systematic variation in the dependent variable?” (Wilson et al., 2010, p. 53)

b. Hva betegnes som ekstern validitet?

“This term refers to the robustness of a phenomenon: the extent to which a causal relationship, once identified in a particular setting with particular research

participants, can safely be generalized to other times, places, and people.” (Wilson et al., 2010, p. 53)

c. Vurder den interne og eksterne validiteten i dette eksperiment.

Relevant pensum (in addition to Wilson et al., 2010):

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the Future: Three Unsuccessful Attempts to Replicate Bem's "Retroactive Facilitation of Recall" Effect. *PLoS ONE*, 7(3), e33423. <http://doi.org/10.1371/journal.pone.0033423>

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15(4), 371–379. <http://doi.org/10.1037/a0025172>

*Minimum (= grade E level for this specific question): Students are expected **to name and describe** at least two factors that might be relevant here. These factors can be in support of or questioning the validity of this experiment. (LeBel & Peters discuss the integrity of measurement instruments and experimental and methodological procedures, Ritchie et al. present attempts to replicate Bem (2011) Experiment 9 and refer to statistical and methodological concerns.)*

*Better answers present more factors and also include **elaborated explanations and arguments for why** a factor is relevant.*