Saker att bestämma om variablerna:

* Hur ska vi göra med de elever som endast har en förälder? Imputera eller endast flytta över värde från förälder 1. 15% saknar data från en förälder 2.
  + Gör analysen med imputerade data och jämför med om man kör analysen utan de imputerade.
* Hur ska vi göra med de enkäter där det finns dubbletter? – Lisa och Jenny: Om det inte finns någon anteckning om orsak till dubbletten så tar vi den enkät som kom in först.
* Hur ska vi göra med “vet ej”-svar I enkäterna? – Förslag Lisa: Koda som NA och imputera. Rapportera antalet som svarade “vet ej” I artikeln
* I VABS: Elev 425, fråga ” Sammanfattande bedomning: Laser och skriver pa samma niva som jamnariga.” läraren har skattat en 2:a med kommentaren ”I den skolformen eleven går i Ja!”- Hur ska vi göra med den? Ändra? Ta bort? Ha kvar?
* I VABS: Elev 106: fråga ” 21. Laser och forstar texter som ar anpassade for ungdomar i aldern 13-15 ar (t.ex. ungdomsbocker, enklare vuxenbocker, tidningsartiklar). (23) 22. Laser och forstar texter som ar anpassade for ungdomar i aldern 16 ar och aldre (t.ex. vuxenlitteratur, facklitteratur, tidningsartiklar). (25)” läraren har skattat en 1:a med kommentaren ”VID HÖGLÄSNING!” – Hur ska vi göra med detta? Läraren har skattat 0 på frågan om eleven läser enklare böcker.
* Jag funderar på om vi ska göra en komposit av våra utbildningsvariabler och använda. Jag har testat modellen och den blir ytterst lik om man endast har med mammans utbildning. Kan inte detta stärka vår variabel? Problemet är hur denna sammanräkning ska ske? Ev. summerar vi endast variablerna för att inte krångla till det för oss.

Bestämt i SEM artikeln:

* Elever med 0 i poäng tas med i analysen
* Val av faktorer i SEMen. 3-4 faktorer per latent variabel. För den latenta variabeln fonologi där vi har flest variabler att välja mellan valde jag de 4 med minst saknad data.
  + Fonologi: letter\_sum, fonem\_sum, fonemsynthesis\_sum, rhyme\_sum
  + Reading: OS64, OLAF, decode\_short\_words, decode\_complex\_words
  + Developmental level: ID\_level, raven (har inga fler variabler att välja mellan)
  + SES: education\_mother, education\_father, occupation\_mother, occupation\_father
* Alla faktorer korrelerade >.3 med de andra faktorerna i samma latenta variabel förutom occupation\_father\*education\_mother (-0,27) men eftersom den var så nära så får den vara med ändå.
* Använder raven som en predicerande variabel snarare än en latent variabel tillsammans med ID-nivå eftersom det endast krånglar till modellen utan att ge något extra.
* Saknad data, hantering:
  + Variabeln education\_mother sakar 9 datapunkter, fem har ersatts av faderns utbildning, 4 av föräldrahemsmammans utbildning
  + OLAF\_sum saknar 1 datapunkt och Decode\_short\_words saknar 7 datapunkter (~5,1%) – dessa punkter imputeras m.h.a. mice()